# INDUSTRIAL MATHEMATICS INSTITUTE

2001:09


# Nonlinear methods of approximation


V.N. Temlyakov



IMI
Preprint Series

Department of Mathematics
University of South Carolina

# Report Documentation Page

*Form Approved*
*OMB No. 0704-0188*



| | | |
|---|---|---|
| 1. REPORT DATE<br>**2001** | 2. REPORT TYPE | 3. DATES COVERED<br>**00-00-2001 to 00-00-2001** |
| 4. TITLE AND SUBTITLE<br>**Nonlinear Methods of Approximation** | | 5a. CONTRACT NUMBER |
| | | 5b. GRANT NUMBER |
| | | 5c. PROGRAM ELEMENT NUMBER |
| 6. AUTHOR(S) | | 5d. PROJECT NUMBER |
| | | 5e. TASK NUMBER |
| | | 5f. WORK UNIT NUMBER |
| 7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)<br>**University of South Carolina,Department of Mathematics,Columbia,SC,29208** | | 8. PERFORMING ORGANIZATION REPORT NUMBER |
| 9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) | | 10. SPONSOR/MONITOR'S ACRONYM(S) |
| | | 11. SPONSOR/MONITOR'S REPORT NUMBER(S) |
| 12. DISTRIBUTION/AVAILABILITY STATEMENT<br>**Approved for public release; distribution unlimited** | | |
| 13. SUPPLEMENTARY NOTES | | |
| 14. ABSTRACT | | |
| 15. SUBJECT TERMS | | |

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| a REPORT<br>**unclassified** | b ABSTRACT<br>**unclassified** | c THIS PAGE<br>**unclassified** | **Same as Report (SAR)** | **58** | |

**Standard Form 298 (Rev. 8-98)**
Prescribed by ANSI Std Z39-18

# Nonlinear Methods of Approximation[1]

V.N.Temlyakov

University of South Carolina, Columbia, SC 29208, USA

Abstract. Our main interest in this paper is nonlinear approximation. The basic idea behind nonlinear approximation is that the elements used in the approximation do not come from a fixed linear space but are allowed to depend on the function being approximated. While the scope of this paper is mostly theoretical, we should note that this form of approximation appears in many numerical applications such as adaptive PDE solvers, compression of images and signals, statistical classification, and so on. The standard problem in this regard is the problem of $m$-term approximation where one fixes a basis and looks to approximate a target function by a linear combination of $m$ terms of the basis. When the basis is a wavelet basis or a basis of other waveforms, then this type of approximation is the starting point for compression algorithms. We are interested in the quantitative aspects of this type of approximation. Namely, we want to understand the properties (usually smoothness) of the function which govern its rate of approximation in some given norm (or metric). We are also interested in stable algorithms for finding good or near best approximations using $m$ terms. Some of our earlier work has introduced and analyzed such algorithms. More recently, there has emerged another more complicated form of nonlinear approximation which we call highly nonlinear approximation. It takes many forms but has the basic ingredient that a basis is replaced by a larger system of functions that is usually redundant. Some types of approximation that fall into this general category are mathematical frames, adaptive pursuit (or greedy algorithms) and adaptive basis selection. Redundancy on the one hand offers much promise for greater efficiency in terms of approximation rate, but on the other hand gives rise to highly nontrivial theoretical and practical problems. With this motivation, our recent work and the current activity focuses on nonlinear approximation both in the classical form of m-term approximation (where several important problems remain unsolved) and in the form of highly nonlinear approximation where a theory is only now emerging.

CONTENTS



[1]This research was supported by the National Science Foundation Grant DMS 9970326 and by ONR Grant N00014-91-J1343

## 1. Introduction

We introduce some notations and orient the reader on the topics that we will be discussing in this paper. We begin our discussion in this section by the case where approximation takes place in a Banach space $X$ equipped with a norm $\|\cdot\| := \|\cdot\|_X$. We formulate our approximation problem in the following general way. We say a set of functions $\mathcal{D}$ from $X$ is a *dictionary* if each $g \in X$ has norm one ($\|g\|_X = 1$) and the closure of Span $\mathcal{D}$ coincides with the whole $X$. We let $\Sigma_m(\mathcal{D})$ denote the collection of all functions (elements) in $X$ which can be expressed as a linear combination of at most $m$ elements of $\mathcal{D}$. Thus each function $s \in \Sigma_m(\mathcal{D})$ can be written in the form

$$s = \sum_{g\in\Lambda} c_g g, \quad \Lambda \subset \mathcal{D}, \quad \#\Lambda \le m,$$

with the $c_g$ are real or complex numbers. In some cases, it may be possible to write an element from $\Sigma_m(\mathcal{D})$ in this form in more than one way. The space $\Sigma_m(\mathcal{D})$ is not linear: the sum of two functions from $\Sigma_m(\mathcal{D})$ is generally not in $\Sigma_m(\mathcal{D})$.

For a function $f \in X$ we define its approximation error

$$\sigma_m(f,\mathcal{D})_X := \inf_{s\in\Sigma_m(\mathcal{D})} \|f - s\|_X,$$

and for a function class $F$

$$\sigma_m(F,\mathcal{D})_X := \sup_{f\in F} \sigma_m(f,\mathcal{D})_X.$$

The classical example of this type of approximation is the case $X = L_p([0,2\pi])$ and $\mathcal{D} = \mathcal{B}$ is an orthogonal basis for $X$. In particular, $\mathcal{B}$ can be taken as the trigonometric system $\mathcal{T} := \{e^{ikx}, \quad k \in \mathbb{Z}\}$ or the Haar system properly normalized. The first results on error estimates in $m$-term approximation showed an advantage of $m$-term approximation over approximation by polynomials of order $m$. R.S. Ismagilov [I] (1974) studied $m$-term trigonometric approximation of individual functions, namely, the Bernoulli kernels

$$F_r(x) = 2\sum_{k=1}^{\infty} k^{-r}\cos(kx - r\pi/2).$$

He proved that

$$\sigma_m(F_2,\mathcal{T})_{L_\infty} \le C_\epsilon m^{-6/5+\epsilon}$$

with arbitrary $\epsilon > 0$. It is known that the best approximation $E_m(\cdot)_{L_\infty}$ by trigonometric polynomials of order $m$ in the $L_\infty$-norm has the asymptotic order $E_m(F_2)_{L_\infty} \asymp 1/m$. Further results in $m$-term trigonometric approximation proved advantage of this type of nonlinear approximation over linear approximation. For many traditional pairs of function class $F$ and orthogonal system $\mathcal{B}$ the orders of $\sigma_m(F,\mathcal{B})_X$ are known now. Investigation of the case $F = B^r_\theta(L_q)$ (standard Besov class), $\mathcal{B} = \mathcal{T}$ and $X = L_p$ was completed in [DT1]. This investigation required new technique (see [DT1] and [KT1]) which uses deep results from finite dimensional geometry. Thus it is an example of interaction between theory of nonlinear $m$-term approximation and contemporary functional analysis. We discuss these results in

detail in Section 6. In Section 6 we also consider a general optimization problem in a spirit of Kolmogorov's widths. For the reader's convenience and for motivation of nonlinear methods we give in Section 5 a brief discussion of optimization settings in the Linear Approximation. Let $\mathbb{D}$ be a collection of dictionaries. The classical example of $\mathbb{D}$ is $\mathbb{O} = \{\text{orthonormal bases on a given domain}\}$. The optimization problem asks to find (if possible) for a given pair of collection of dictionaries $\mathbb{D}$ and function class $F$ a dictionary $\mathcal{D} \in \mathbb{D}$ such that

$$\sigma_m(F, \mathcal{D})_X \asymp \sigma_m(F, \mathbb{D})_X := \inf_{\mathcal{D} \in \mathbb{D}} \sigma_m(F, \mathcal{D})_X.$$

This problem is interesting and important for theoretical investigation and also for practical applications where we often want to have a dictionary $\mathcal{D}$ with certain structure (from a collection $\mathbb{D}$) and do not want to stick to a particular one. In Section 6 we discuss only theoretical aspect of this problem for the classical example of $\mathbb{D} = \mathbb{O}$.

The next problem that we propose to investigate is to find a universal dictionary $\mathcal{D} \in \mathbb{D}$, i.e. the one which is optimal for all $F$ from a given collection $\mathcal{F}$ of function classes.

**Definition 1.1.** *Let two collections $\mathcal{F}$ of function classes and $\mathbb{D}$ of dictionaries be given. We say that $\mathcal{D} \in \mathbb{D}$ is universal for the pair $(\mathcal{F}, \mathbb{D})$ if there exists a constant $C$ which may depend on $\mathcal{F}$, $\mathbb{D}$ and $X$ such that for any $F \in \mathcal{F}$ we have*

$$\sigma_m(F, \mathcal{D})_X \le C\sigma_m(F, \mathbb{D})_X. \tag{1.1}$$

It may happen that for a given pair $(\mathcal{F}, \mathbb{D})$ there is no universal dictionaries. In this case we define the index of universality and look for a dictionary which realizes (in the sense of order) this index. Let $m$ be fixed. Take a dictionary $\mathcal{D} \in \mathbb{D}$ and for a fixed $F \in \mathcal{F}$ find the minimal $N(m, \mathcal{D}, F)$ such that

$$\sigma_{N(m,\mathcal{D},F)}(F, \mathcal{D})_X \le \sigma_m(F, \mathbb{D})_X.$$

We define index of universality by

$$iu(\mathcal{F}, \mathbb{D}, m) := \inf_{\mathcal{D} \in \mathbb{D}} \sup_{F \in \mathcal{F}} \frac{N(m, \mathcal{D}, F)}{m}.$$

This is a new concept in nonlinear approximation. The following observation motivates our interest in this setting. In practice we often do not know the exact smoothness class $F$ where our input function (signal, image) comes from. Instead, we often know that our function comes from a class of certain structure, for instance, anisotropic Sobolev class. This is exactly the situation we are dealing with in the universal dictionary setting. So, if for a collection $\mathcal{F}$ there exists a universal dictionary $\mathcal{D} \in \mathbb{D}$ it is an ideal situation. We can use this universal dictionary $\mathcal{D}$ in all cases and we know that it ajusts automatically to the best smoothness class $F \in \mathcal{F}$ which contains a function under approximation. Next, if a pair $(\mathcal{F}, \mathbb{D})$ does not allow a universal dictionary we have a trade off between universality and accuracy provided by the index of universality. We discuss the universality results in Section 7.

We discussed above best $m$-term approximation with regard to a dictionary $\mathcal{D}$ in a Banach space $X$. The sequence $\{\sigma_m(f,\mathcal{D})_X\}$ gives the lower estimates of accuracy for any sequence of algorithms $A_m$ that map $X$ into $\Sigma_m(\mathcal{D})$, where, as above $\Sigma_m(\mathcal{D})$ is the set of all functions in $X$ which can be expressed as a linear combination of at most $m$ elements from $\mathcal{D}$. Thus, the sequences $\{\sigma_m(f,\mathcal{D})_X\}$ and $\{\sigma_m(F,\mathcal{D})_X\}$ may serve as the target accuracies in constructing approximating algorithms $A_m$. It is clear that the best algorithm (if exists) gives the error

$$\|f - A_m(f,\mathcal{D})\|_X = \sigma_m(f,\mathcal{D})_X. \tag{1.2}$$

We call an algorithm $A_m$ near best or near best for individual functions if

$$\|f - A_m(f,\mathcal{D})\|_X \le C(\mathcal{D},X)\sigma_m(f,\mathcal{D})_X \tag{1.3}$$

for all $f \in X$. Similarly, we say that $A_m$ is near best for a function class $F$ if we have for any $f \in F$

$$\|f - A_m(f,\mathcal{D})\|_X \le C(F,\mathcal{D},X)\sigma_m(F,\mathcal{D})_X. \tag{1.4}$$

It is clear that an algorithm $A_m$ satisfying (1.3) is excellent from the point of veiw of accuracy: it provides near best approximation for every individual function and, therefore, for any function class. The property (1.4) is weaker than (1.3) but still is very good. We present in Section 2 some results on linear approximation of individual functions and function classes. The corresponding results for nonlinear approximation with regard to a basis are discussed in Section 3. Let a Banach space $X$ with a basis $\Psi = \{\psi_k\}_{k=1}^{\infty}$, $\|\psi_k\| = 1$, $k = 1,2,\dots$, be given. We consider the following theoretical greedy algorithm that we call Thresholding Greedy Algorithm (TGA). For a given element $f \in X$ we consider the expansion

$$f = \sum_{k=1}^{\infty} c_k(f)\psi_k.$$

Let an element $f \in X$ be given. We call a permutation $\rho$, $\rho(j) = k_j$, $j = 1,2,\dots$, of the positive integers decreasing and write $\rho \in D(f)$ if

$$|c_{k_1}(f)| \ge |c_{k_2}(f)| \ge \dots \quad .$$

In the case of strict inequalities here $D(f)$ consists of only one permutation. We define the $m$-th greedy approximant of $f$ with regard to the basis $\Psi$ corresponding to a permutation $\rho \in D(f)$ by formula

$$G_m(f,\Psi) := G_m^X(f,\Psi) := G_m^X(f,\Psi,\rho) := \sum_{j=1}^{m} c_{k_j}(f)\psi_{k_j}.$$

It is a simple algorithm which describes theoretical scheme (it is not computationally ready) for $m$-term approximation of an element $f$.

We have discussed above the general optimization setting to find a good basis for nonlinear approximation. On the base of this discussion we suggest a three step strategy to find a good basis (dictionary) for nonlinear $m$-term approximation.

The first step consists of solving an optimization problem for a given function class $F$, when we optimize over a collection $\mathbb{D}$ of bases (dictionaries). The second step is devoted to finding a universal basis (dictionary) $\mathcal{D}_u \in \mathbb{D}$ for a given pair $(\mathcal{F}, \mathbb{D})$ of collections: $\mathcal{F}$ of function classes and $\mathbb{D}$ of bases (dictionaries). The third step deals with constructing a theoretical algorithm that realizes near best $m$-term approximation with regard to $\mathcal{D}_u$ for function classes from $\mathcal{F}$. We worked this strategy out in the model case of anisotropic function classes and the set of orthogonal bases. The results are positive. We constructed a natural tensor-product-wavelet type basis and proved that it is universal. Moreover, we proved that Thresholding Greedy Algorithm realizes near best $m$-term approximation with regard to this basis for all anisotropic function classes. We discuss these results in Section 7.

It is also very important to find analogs of $G_m^X(\cdot, \Psi)$ in the case of general dictionary $\mathcal{D}$ and to study their efficiency. We start this discussion with confining ourselves to Hilbert spaces. We define first the Pure Greedy Algorithm (PGA) in Hilbert space $H$. We describe this algorithm for a general dictionary $\mathcal{D}$. If $f \in H$, we let $g(f) \in \mathcal{D}$ be an element from $\mathcal{D}$ which maximizes $|\langle f, g\rangle|$. We shall assume for simplicity that such a maximizer exists; if not suitable modifications are necessary (see Weak Greedy Algorithm below) in the algorithm that follows. We define

$$G(f, \mathcal{D}) := \langle f, g(f)\rangle g(f)$$

and

$$R(f, \mathcal{D}) := f - G(f, \mathcal{D}).$$

**Pure Greedy Algorithm (PGA).** *We define $R_0(f, \mathcal{D}) := f$ and $G_0(f, \mathcal{D}) := 0$. Then, for each $m \geq 1$, we inductively define*

$$G_m(f, \mathcal{D}) := G_{m-1}(f, \mathcal{D}) + G(R_{m-1}(f, \mathcal{D}), \mathcal{D})$$

$$R_m(f, \mathcal{D}) := f - G_m(f, \mathcal{D}) = R(R_{m-1}(f, \mathcal{D}), \mathcal{D}).$$

In Section 8 we consider the problem of efficiency of Pure Greedy Algorithms with regard to general dictionaries in Hilbert space.

**Remark 1.1.** In this paper, we study only theoretical aspects of the efficiency of $m$-term approximation and possible ways to realize this efficiency. The above defined "greedy algorithm" gives a procedure to construct an approximant which turns out to be a good approximant. The procedure of constructing a greedy approximant is not a numerical algorithm ready for computational implementation. Therefore it would be more precise to call this procedure a "theoretical greedy algorithm" or "stepwise optimizing process". Keeping this remark in mind we, however, use term "greedy algorithm" in this paper because it has been used in previous papers and has become a standard name for procedures like the above and for more general procedures of this type (see for instance [D], [DT2]). Following [DDGS] we call an algorithm "incremental" if at step $m$ we add at most one more element $\varphi_m \in \mathcal{D}$ and approximate by linear combination $c_1\varphi_1 + \cdots + c_m\varphi_m$. We use the term "greedy type" for an incremental algorithm with $\varphi_m$ chosen to maximize a given functional $F(f_{m-1}, g)$ over $g \in \mathcal{D}$ with $f_{m-1}$ is a residual after the $(m-1)$th step of the algorithm. The form of $F(\cdot, \cdot)$ determines the kind of greedy algorithm. We use

the term "weak greedy" for an incremental algorithm with $\varphi_m$ satisfying a weaker condition than maximizing the given functional. For instance,

$$F(f^\tau_{m-1},\varphi_m) \ge t_m \sup_{g\in\mathcal{D}} F(f^\tau_{m-1},g), \quad 0\le t_m\le 1.$$

The sequence $\tau := \{t_k\}_{k=1}^\infty$ is called the "weakness" sequence.

We begin with two definitions in a spirit of inequalities (1.3) and (1.4).

**Definition 1.2.** *We call a dictionary $\mathcal{D}$ greedy dictionary for a Hilbert space $H$ if for any $f\in H$ and any realization of Pure Greedy Algorithm we have*

$$\|f-G_m(f,\mathcal{D})\| \le C(\mathcal{D},H)\sigma_m(f,\mathcal{D}).$$

**Definition 1.3.** *Let $r>0$ be given. We call a dictionary $\mathcal{D}$ $r$-greedy dictionary for $H$ if $\mathcal{D}$ posses the property (G): for any $f\in H$ such that*

$$\sigma_m(f,\mathcal{D}) \le m^{-r}, \quad m=1,2,\dots,$$

*we have*

$$\|f-G_m(f,\mathcal{D})\| \le C(r,\mathcal{D})m^{-r}, \quad m=1,2,\dots.$$

A simple example of greedy dictionary is an orthonormal basis for $H$. There is a nontrivial classical example of greedy dictionary. Let $\Pi$ be a set of functions from $L_2([0,1]^2)$ of the form $u(x_1)v(x_2)$ with unit $L_2$-norm. Then for this dictionary and $H=L_2([0,1]^2)$ we have for each $f\in H$

$$\|f-G_m(f,\Pi)\| = \sigma_m(f,\Pi).$$

This result and related results will be discussed in Section 11. We will discuss in Section 8 the general setting for Pure Greedy Algorithm and modifications of PGA some of which we define now. For other modifications see [DT2] and [DDGS].

Let a sequence $\tau=\{t_k\}_{k=1}^\infty$, $0\le t_k\le 1$, be given. Following [T20] we define Weak Greedy Algorithm.

**Weak Greedy Algorithm (WGA).** *We define $f_0^\tau := f$. Then for each $m\ge 1$, we inductively define:*

*1). $\varphi_m^\tau\in\mathcal{D}$ is any satisfying*

$$|\langle f^\tau_{m-1},\varphi^\tau_m\rangle| \ge t_m \sup_{g\in\mathcal{D}} |\langle f^\tau_{m-1},g\rangle|;$$

*2).*

$$f^\tau_m := f^\tau_{m-1} - \langle f^\tau_{m-1},\varphi^\tau_m\rangle\varphi^\tau_m;$$

*3).*

$$G^\tau_m(f,\mathcal{D}) := \sum_{j=1}^m \langle f^\tau_{j-1},\varphi^\tau_j\rangle\varphi^\tau_j.$$

We note that in a particular case $t_k=t$, $k=1,2,\dots$, this algorithm was considered in [J1]. We present convergence results and error estimates for PGA and WGA in Section 8.

Much less is known about greedy algorithms in the case of Banach space $X$. We discuss here two versions of generalization of PGA from Hilbert space $H$ to Banach space $X$. The first one is a straightforward generalization of PGA. We call it Pure Greedy Algorithm or $X$-Greedy Algorithm when we want to point out a Banach space. For a given $X$ and $\mathcal{D}$ we define $G(f, \mathcal{D}, X) := \alpha(f)g(f)$ where $\alpha(f) \in \mathbb{R}$ and $g(f) \in \mathcal{D}$ satisfy (we assume existence) the relation

$$\min_{\alpha \in \mathbb{R}, g \in \mathcal{D}} \|f - \alpha g\| = \|f - \alpha(f)g(f)\|.$$

**$X$-Greedy Algorithm.** *We define $R_0(f, \mathcal{D}, X) := f$ and $G_0(f, \mathcal{D}, X) := 0$. Then, for each $m \geq 1$, we inductively define*

$$R_m(f) := R_m(f, \mathcal{D}, X) := R_{m-1}(f) - G(R_{m-1}(f), \mathcal{D}, X)$$

$$G_m(f, \mathcal{D}, X) := G_{m-1}(f, \mathcal{D}, X) + G(R_{m-1}(f), \mathcal{D}, X).$$

The second version of PGA in Banach space is based on the concept of peak functional (norming functional). We call it Dual Greedy Algorithm (DGA). Let a dictionary $\mathcal{D}$ in $X$ be given. Take an element $f \in X$ and find a peak functional $F_f$, i.e. a functional such that $\|F_f\|_{X'} = 1$ and $F_f(f) = \|f\|_X$. The existence of such a functional follows from the Hahn-Banach theorem. Now the basic step of PGA is modified to the following. Assume that there exists $g_f \in \mathcal{D}$ such that

$$|F_f(g_f)| = \max_{g \in \mathcal{D}} |F_f(g)|.$$

We take this $g_f$ and solve one more optimization problem: find a number $a$ such that

$$\|f - ag_f\|_X = \min_b \|f - bg_f\|_X.$$

We put

$$G^D(f, \mathcal{D}) := ag_f, \quad R^D(f, \mathcal{D}) := f - ag_f.$$

Repeating this step $m$ times we get $G^D_m(f, \mathcal{D})$ as an approximant and $R^D_m(f, \mathcal{D})$ as a residual. Some results on greedy algorithms in Banach spaces are presented in Section 9.

In Section 10 we discuss some results on how the entropy numbers can be used in estimating from below the quantities $\{\sigma_m(F, \Psi)\}$. The idea of estimating the Kolmogorov widths from below using the entropy numbers is well known (see [L], [C], [Pi]). We used this idea in [T16] for estimating nonlinear best $m$-term approximation. We proved that for good systems $\Psi$ the estimate

$$\epsilon_n(F, X) \gg n^{-a}(\log n)^b, \quad a > 0, b \in \mathbb{R},$$

for the entropy numbers implies the same estimate for best $m$-term approximation:

$$\sigma_m(F, \Psi)_X \gg m^{-a}(\log m)^b \quad .$$

See Section 10 for more detail.

We mention two survey papers [Bab] and [D] where one can find detailed discussion of numerical applications.

## 2. Approximation by Linear Methods. Individual functions

Let us consider a Banach space $X$ with a basis $\Psi = \{\psi_k\}_{k=1}^{\infty}$, $\|\psi_k\| = 1$, $k = 1,2,\dots$. For a given element $f \in X$ we consider the expansion

$$f = \sum_{k=1}^{\infty} c_k(f)\psi_k$$

and the correspoding partial sums

$$S_n(f,\Psi) := \sum_{k=1}^{n} c_k(f)\psi_k.$$

In order to understand efficiency of approximating by $S_n$ we introduce best approximations with regard to $\operatorname{Span}\{\psi_1,\dots,\psi_n\}$:

$$E_n(f,\Psi)_X := \inf_{a_k} \|f - \sum_{k=1}^{n} a_k\psi_k\|_X.$$

It is well known (see [LT]) that for a basis $\Psi$ the operator $S_n$ is bounded as an operator from $X$ to $X$. Therefore, we have for any $f,g \in X$

$$\|S_n(f) - S_n(g)\|_X \le C(X,\Psi)\|f-g\|_X,$$

and for any $f \in X$

$$\|f - S_n(f,\Psi)\|_X \le C(X,\Psi)E_n(f,\Psi)_X.$$

This means that the partial sums method provides near best approximation for any individual $f$. Let us consider a classical example of $\Psi = \mathcal{T}$ - the trigonometric system and $X = L_p$, $1 \le p \le \infty$. The basis $\mathcal{T}$ is an orthonormal basis and, therefore, orthoprojector $S_n$ realizes the best approximation in $L_2$. By the Riesz theorem (see [Z]) we know that $\mathcal{T}$ is a basis for $1 < p < \infty$ and thus the Fourier sums realize near best trigonometric approximation in $L_p$, $1 < p < \infty$. It is well known that $\mathcal{T}$ is not a basis for $L_1$ and $L_\infty$. In this case we have the Lebesgue inequality:

$$\|f - S_n(f,\mathcal{T})\|_p \le C\ln(n+2)E_n(f,\mathcal{T})_p, \quad p = 1,\infty.$$

An extra factor $\ln(2+n)$ is a slowly growing to infinity function on $n$ but nonetheless there are different settings where an attempt to get rid of $\ln(2+n)$ was done. We will mention some of them. One can replace the partial sum $S_n(f,\mathcal{T})$ by the de la Vallée-Poussin operator

$$V_n(f,\mathcal{T}) := \frac{1}{n}\sum_{j=n}^{2n-1} S_j(f,\mathcal{T}).$$

It is not an orthoprojector anymore but one has the estimate

$$\|f - V_n(f,\mathcal{T})\|_p \le 4E_n(f,\mathcal{T})_p, \quad p = 1,\infty,$$

that is good if $\{E_n(f,\mathcal{T})_p\}$ does not decrease fast (note that $V_n(f,\mathcal{T})$ is a trigonometric polynomial of degree $2n-1$). The following estimate was obtained by Oskolkov [O2] for $p=\infty$

$$\|f-S_n(f,\mathcal{T})\|_\infty \le C\sum_{j=n}^{2n}\frac{E_j(f,\mathcal{T})_\infty}{j-n+1}.$$

We note also that in the case of $p=\infty$ an extra $\ln(2+n)$ appears not only in the estimates for individual functions as above but also for function classes. We present here some well known results for the Sobolev classes

$$W^r_q := \{f \quad : \quad f^{(r-1)}\text{-absolutely continuous}, \quad \|f^{(r)}\|_q \le 1\}.$$

Kolmogorov proved that

$$\sup_{f\in W^r_\infty}\|f-S_n(f,\mathcal{T})\|_\infty = \frac{4}{\pi^2}(\ln n)n^{-r}+O(n^{-r}).$$

Favard, Akhiezer and Krein (see [Tim]) proved the equality

$$\sup_{f\in W^r_\infty}E_n(f,\mathcal{T})_\infty = K_r(n+1)^{-r},$$

with $K_r$ is a number depending on the number $r$.

We discuss an interplay between approximation of individual functions and function classes. In this section we discuss certain aspects of the following question. Suppose that $F$ is a function class and $\{\delta_n(F)\}_{n=1}^\infty$ is a corresponding sequence of extremal quantities. In this section we take $\delta_n(F):=\sup_{f\in F}\delta_n(f)$ to be the supremum $e_n(F)$ or $E_n(F)$ of the best approximation in the uniform norm of functions in $F$ by algebraic $e_n(\cdot)$ or trigonometric $E_n(\cdot)$ polynomials of order $n$. In Section 5 we will consider the case $\delta_n(F)=d_n(F)$ - the sequence of the Kolmogorov widths of the class $F$. We discuss the question of the extent to which the sequence $\{\delta_n(F)\}_{n=1}^\infty$, which is connected with the whole function class $F$, characterizes the corresponding properties of individual functions in $F$. In this section we discuss the question of the existence in $F$ of a function $f$ such that

$$\lim_{n\to\infty}\delta_n(f)/\delta_n(F)=1.$$

The first result in this direction is apparently due to Lebesgue. In [Le] he proved the equality

$$\sup_{\|f\|_\infty\le M}E_n(f)_\infty = M,$$

where sup is taken over continuous functions. This equality in combination with the Weierstrass theorem shows that in the class of all continuous functions bounded by the number $M$ there is no asymptotically extremal function.

Let us make a historical remark due to Nikol'skii (see [N2]). S.N. Bernstein discussed the role of function classes in constructive approximation in the opening session of his seminar in Approximation Theory (Moscow, Spring 1945). His general attitude to the role of studying the sequences of $E_n(F):=\sup_{f\in F}E_n(f)$ for a given

function class $F$ was skeptical. One of his arguments was that the sequence $\{E_n(F)\}$ may not reflect the behavior of $\{E_n(f)\}$ for any individual $f \in F$, because usually the extreme function that realizes $\sup_{f\in F} E_n(f)$ depends on $n$. He formulated a problem of studying

$$\sup_{f\in F} \limsup_{n\to\infty} \frac{E_n(f)}{E_n(F)} \quad \text{and} \quad \sup_{f\in F} \liminf_{n\to\infty} \frac{E_n(f)}{E_n(F)}$$

and their analogs for approximation by algebraic polynomials for some function classes. In particular he thought that the function $|x|$ is an extremal function in the sense of the above quantities in the class $Lip_1 1$ for approximation by algebraic polynomials in the uniform norm. However, it turned out not to be the case. S. M. Nikol'skii [N2] proved in 1946 that for $W^r_\infty$ classes there is a function $f \in W^r_\infty$ such that

$$\limsup_{n\to\infty} E_n(f)/E_n(W^r_\infty) = 1.$$

It was proved in [T1], [T2] that for the class $W^r_\infty$ there exists a function $f \in W^r_\infty$ such that

$$\lim_{n\to\infty} E_n(f)/E_n(W^r_\infty) = 1. \tag{2.1}$$

Further results and some generalizations are obtained in [T3], [T5]. It is interesting to compare the above result (2.1) with the following result of Oskolkov [O1]

$$\max_{f\in W^1_\infty} \liminf_{n\to\infty} (\|f - S_n(f,\mathcal{T})\|_\infty / \sup_{f\in W^1_\infty} \|f - S_n(f,\mathcal{T})\|_\infty) = 1/2.$$

**Open problem 2.1.** Is it possible to extend (2.1) from $W^r_\infty$ to $W^r H^\omega$ with arbitrary modulus of continuity $\omega$? We define here

$$W^r H^\omega := \{f \quad : \quad |f^{(r)}(x) - f^{(r)}(y)| \le \omega(|x-y|), \quad x, y \in \mathbb{T}\}.$$

## 3. Greedy Approximation with regard to bases

**3.1 Greedy Bases.** We will study the algorithms $G_m(f,\Psi,\rho)$ defined in the Introduction. In order to understand the efficiency of this algorithm we compare its accuracy with the best possible $\sigma_m(f,\Psi)$ when an approximant is a linear combination of $m$ terms from $\Psi$. The best we can achieve with the algorithm $G_m$ is

$$\|f - G_m(f,\Psi,\rho)\| = \sigma_m(f,\Psi),$$

or a little weaker

$$\|f - G_m(f,\Psi,\rho)\| \le G\sigma_m(f,\Psi) \tag{3.1}$$

for all elements $f \in X$ with a constant $G = C(X,\Psi)$ independent of $f$ and $m$.

**Definition 3.1.** *We call a basis $\Psi$ greedy basis if for every $f \in X$ there exists a permutation $\rho \in D(f)$ such that (3.1) holds.*

The following proposition has been proved in [KoT1].

**Proposition 3.1.** *If $\Psi$ is a greedy basis then (3.1) holds for any permutation $\rho \in D(f)$.*

We will discuss two the most interesting cases of basis $\Psi$: the Haar basis $\mathcal{H}$ as a representative of wavelet type bases and the trigonometric system $\mathcal{T}$ as a representative of uniformly bounded orthonormal bases.

Denote $\mathcal{H}_p := \{H_k^p\}_{k=1}^\infty$ the Haar basis on $[0,1)$ normalized in $L_p(0,1)$: $H_1^p = 1$ on $[0,1)$ and for $k = 2^n + l$, $n = 0,1,\dots$, $l = 1,2,\dots,2^n$

$$H_k^p = \begin{cases} 2^{n/p}, & x \in [(2l-2)2^{-n-1}, (2l-1)2^{-n-1}) \\ -2^{n/p}, & x \in [(2l-1)2^{-n-1}, 2l2^{-n-1}) \\ 0, & \text{otherwise.} \end{cases}$$

Denote by $\mathcal{T} := \{e^{ikx}\}_{k\in\mathbb{Z}}$ the univariate trigonometric system in the complex form and denote by $\mathcal{T}^d := \mathcal{T} \times \cdots \times \mathcal{T}$ the multivariate trigonometric system.

The following theorem (see [T14]) establishes existence of greedy bases for $L_p(0,1)$, $1 < p < \infty$.

**Theorem 3.1.** *Let $1 < p < \infty$ and a basis $\Psi$ be $L_p$-equivalent to the Haar basis $\mathcal{H}_p$. Then for any $f \in L_p(0,1)$ and any $\rho \in D(f)$ we have*

$$\|f - G_m(f,\Psi,\rho)\|_{L_p} \le C(p,\Psi)\sigma_m(f,\Psi)_{L_p}$$

*with a constant $C(p,\Psi)$ independent of $f$, $\rho$, and $m$.*

We use in this theorem the following definition of the $L_p$-equivalence. We say that $\Psi = \{\psi_k\}_{k=1}^\infty$ is $L_p$-equivalent to $\mathcal{H}_p = \{H_k^p\}_{k=1}^\infty$ if for any finite set $\Lambda$ and any coefficients $c_k$, $k \in \Lambda$, we have

$$C_1(p,\Psi)\|\sum_{k\in\Lambda} c_k H_k^p\|_{L_p} \le \|\sum_{k\in\Lambda} c_k\psi_k\|_{L_p} \le C_2(p,\Psi)\|\sum_{k\in\Lambda} c_k H_k^p\|_{L_p}$$

with two positive constants $C_1(p,\Psi), C_2(p,\Psi)$ which may depend on $p$ and $\Psi$. For sufficient conditions on $\Psi$ to be $L_p$-equivalent to $\mathcal{H}_p$ see [FJ] and [DKT].

Thus each basis $\Psi$ which is $L_p$-equivalent to the univariate Haar basis $\mathcal{H}_p$ is a greedy basis for $L_p(0,1)$, $1 < p < \infty$. We note that in the case of Hilbert space each orthonormal basis is a greedy basis with a constant $G = 1$ (see (3.1)).

We give now the definitions of unconditional and democratic bases.

**Definition 3.2.** *A basis $\Psi = \{\psi_k\}_{k=1}^\infty$ of a Banach space $X$ is said to be unconditional if for every choice of signs $\theta = \{\theta_k\}_{k=1}^\infty$, $\theta_k = 1$ or $-1$, $k = 1,2,\dots$, the linear operator $M_\theta$ defined by*

$$M_\theta(\sum_{k=1}^\infty a_k\psi_k) = \sum_{k=1}^\infty a_k\theta_k\psi_k$$

*is a bounded operator from $X$ into $X$.*

**Definition 3.3.** *We say that a basis $\Psi = \{\psi_k\}_{k=1}^\infty$ is a democratic basis if for any two finite sets of indices $P$ and $Q$ with the same cardinality $\#P = \#Q$ we have*

$$\|\sum_{k\in P}\psi_k\| \le D\|\sum_{k\in Q}\psi_k\|$$

*with a constant $D := D(X,\Psi)$ independent of $P$ and $Q$.*

We proved in [KoT1] the following theorem.

**Theorem 3.2.** *A basis is greedy if and only if it is unconditional and democratic.*

We remark that Definition 3.1 of greedy basis for a Banach space is an analog of Definition 1.2 (see Introduction) of greedy dictionary for a Hilbert space. Let us give an analog to Definition 1.3 of $r$-greedy dictionary.

**Definition 3.4.** *We call a basis $\Psi$ $r$-greedy basis for a Banach space $X$ if for each $f \in X$ such that*

$$\sigma_m(f, \Psi)_X \le m^{-r}, \quad m = 1, 2, \dots,$$

*we have for every $\rho \in D(f)$*

$$\|f - G_m(f, \Psi, \rho)\| \le C(r, \Psi) m^{-r}, \quad m = 1, 2, \dots.$$

We construct the following example now.

**Example 3.1.** *There exist a Banach space $X$ and a basis $\Psi$ such that $\Psi$ is a $r$-greedy basis for $X$ for any $r > 0$ and $\Psi$ is not an unconditional basis.*

*Proof.* We use the construction from [KoT1]. Let $X$ be the set of all real sequences $x = (x_1, x_2, \dots) \in l_2$ such that

$$||x||' = \sup_{N \in \mathbf{N}} \left| \sum_{n=1}^{N} x_n / \sqrt{n} \right|$$

is finite. Clearly, $X$ equipped with the norm

$$|| \cdot || = \max(|| \cdot ||_{l_2}, || \cdot ||')$$

is a Banach space. Let $\psi_k \in X$, $k = 1, 2, \dots$, be defined as

$$(\psi_k)_n = \begin{cases} 1, & n = k, \\ 0, & n \ne k. \end{cases}$$

We take any $r > 0$ and prove that $\Psi$ is $r$-greedy basis for $X$. Indeed, the assumption $\sigma_m(f, \Psi)_X \le m^{-r}$ implies $\sigma_m(f, \Psi)_{l_2} \le m^{-r}$ and, therefore,

$$\|f - G_m(f, \Psi)\|_{l_2} \le m^{-r}.$$

Let us prove a similar estimate for $\| \cdot \|'$. Let

$$G_m(f, \Psi) = \sum_{k \in \Lambda_m} c_k(f) \psi_k.$$

Denote $Q_m(N) := [1, N] \setminus \Lambda_m$. Then

$$\|f - G_m(f, \Psi)\|' = \sup_N | \sum_{k \in Q_m(N)} c_k(f) k^{-1/2}| \le \sum_{k=1}^{\infty} k^{-1/2} (m + k)^{-r-1/2} \ll m^{-r}.$$

This proves that $\Psi$ is a $r$-greedy basis for $X$. It is proved in [KoT1] that $\Psi$ is not unconditional.

**3.2. The Trigonometric System.** Let us consider nonlinear approximation with regard to the trigonometric system $\mathcal{T}^d$. The existence of best $m$-term trigonometric approximation was proved in [Ba] (see also [T19]). The method $G_m(f) := G_m(f, \mathcal{T}^d)$ has one more advantage over the traditional approximation by trigonometric polynomials in the case of approximation of functions of several variables. In this case ($d > 1$) there is no natural order of trigonometric system and the use of $G_m$ allows us to avoid the problem of finding natural subspaces of trigonometric polynomials for approximation purposes. We proved in [T19] the following inequality.

**Theorem 3.3.** *For each $f \in L_p(\mathbb{T}^d)$ we have*

$$\|f - G_m(f)\|_p \le (1 + 3m^{h(p)})\sigma_m(f)_p, \quad 1 \le p \le \infty,$$

*where $h(p) := |1/2 - 1/p|$.*

**Remark 3.1.** *For all $1 \le p \le \infty$*

$$\|G_m(f)\|_p \le m^{h(p)}\|f\|_p.$$

**Remark 3.2.** *There is a positive absolute constant $C$ such that for each $m$ and $1 \le p \le \infty$ there exists a function $f \neq 0$ with the property*

$$\|G_m(f)\|_p \ge Cm^{h(p)}\|f\|_p. \tag{3.2}$$

The above results show that the trigonometric system is not a greedy basis for $L_p$, $p \neq 2$. This leads to a natural attempt to consider some other algorithms that may have some advantages over TGA in the case of $\mathcal{T}$. We discuss here the performance of WCGA (see Section 9) with regard to $\mathcal{T}$.

Let us compare the rate of approximation of TGA and WCGA for the class $A := A(\mathcal{RT})$ where $\mathcal{RT}$ denotes the real trigonometric system $1/2, \sin x, \cos x, \dots$. We need to switch to this system from the complex trigonometric system because the algorithm WCGA is defined for the real Banach space. We note that the system $\mathcal{RT}$ is not normalized in $L_p$ but quasinormalized: $C_1 \le \|t\|_p \le C_2$ for any $t \in \mathcal{RT}$ with absolute constants $C_1$, $C_2$, $1 \le p \le \infty$. It is sufficient for application of general methods developed in Section 9. For a sequence $\tau := \{t_k\}$ with $t_k = t$, $k = 1, 2, \dots,$ we replace $\tau$ by $t$ in the notation. Theorem 9.1 and (9.6) imply the following result.

**Theorem 3.4.** *Let $0 < t \le 1$. For $f \in A$ we have*

$$\|f - G^{c,t}_m(f, \mathcal{RT})\|_p \le C(p,t)m^{-1/2}, \quad 2 \le p < \infty. \tag{3.3}$$

This estimate and Theorem 3.3 imply that for $f \in A$ we have

$$\|f - G_m(f, \mathcal{RT})\|_p \le C(p,t)m^{-1/p}, \quad 2 \le p < \infty, \tag{3.4}$$

what is weaker than (3.3). It is proved in [DKTe] that (3.4) can not be improved. Thus the WCGA works better than the TGA for the class $A$. We note that the restriction $p < \infty$ in (3.3) is important. We give now a lower estimate for $m$-term approximation in $L_\infty$.

**Proposition 3.2.** *For a given $m$ define*

$$f := \sum_{k=0}^{2m} \cos 3^k x.$$

*Then we have*

$$\sigma_m(f, \mathcal{T})_\infty \geq m/4.$$

*Proof.* Consider the Riesz product

$$\Phi_0(x) := \prod_{j\in[0,2m]} (1 + \cos 3^j x) - 1.$$

This function has nonzero Fourier coefficients only with frequences of the form

$$k(s) = \sum_{j=0}^{j(s)} s_j 3^j, \quad s = (s_0, \dots, s_{2m}),$$

with $0 \leq j(s) \leq 2m$, $s_j = -1, 0, 1$ for $j < j(s)$, $s_{j(s)} = 1$, and $s_j = 0$ for $j(s) < j \leq 2m$. It is clear that $k(s)$ is uniquely defined by $s$. Take any polynomial of the form

$$t(x) = \sum_{k\in\Lambda} a_k \cos kx, \quad \#\Lambda = m.$$

Then for each $k \in \Lambda$ we look for an $s$ such that $k = k(s)$. If we do not find such an $s$ we have

$$\langle \cos kx, \Phi_0 \rangle = 0.$$

For those $s$ that were found to satisfy $k(s) = k$, $k \in \Lambda$, we form a set $J$ consisting of all $j(s)$ and define the new Riesz product

$$\Phi := \prod_{j\in[0,2m]\setminus J} (1 + \cos 3^j x) - 1.$$

Then we have

$$\langle t, \Phi \rangle = 0$$

and

$$m + 1 \leq \langle f - t, \Phi \rangle \leq \|f - t\|_\infty \|\Phi\|_1 \leq 4\|f - t\|_\infty.$$

This implies

$$\sigma_m(f, \mathcal{T})_\infty \geq m/4.$$

**3.3 Greedy Bases. Direct and Inverse Theorems.** Theorem 3.1 points out the importance of bases $L_p$-equivalent to the Haar basis. We will discuss now necessary and sufficient conditions for $f$ to have a prescribed decay of $\{\sigma_m(f, \Psi)_p\}$ under assumption that $\Psi$ is $L_p$-equivalent to the Haar basis $\mathcal{H}_p$, $1 < p < \infty$. We will express these conditions in terms of coefficients $\{f_n\}$ of the expansion

$$f = \sum_{n=1}^{\infty} f_n \psi_n.$$

The following lemma from [T14] plays the key role in this consideration.

**Lemma 3.1.** *Let a basis $\Psi$ be $L_p$-equivalent to $\mathcal{H}_p$, $1 < p < \infty$. Then for any finite $\Lambda$ and $a \le |c_n| \le b$, $n \in \Lambda$, we have*

$$C_1(p,\Psi)a(\#\Lambda)^{1/p} \le \|\sum_{n\in\Lambda} c_n\psi_n\|_p \le C_2(p,\Psi)b(\#\Lambda)^{1/p}.$$

We formulate a general statement and then consider several important particular examples of rate of decrease of $\{\sigma_m(f,\Psi)_p\}$. We begin by introducing some notations. For a monotonically decreasing to zero sequence $\mathcal{E} = \{\epsilon_k\}_{k=0}^{\infty}$ of positive numbers (we write $\mathcal{E} \in MDP$) we define inductively a sequence $\{N_s\}_{s=0}^{\infty}$ of nonnegative integers:

$$N_0 = 0; \quad N_{s+1} \quad \text{is the smallest satisfying} \tag{3.5}$$

$$\epsilon_{N_{s+1}} < \frac{1}{2}\epsilon_{N_s}; \qquad n_s := N_{s+1} - N_s.$$

We are going to consider the following examples of sequences.

**Example 3.2.** *Take $\epsilon_0 = 1$ and $\epsilon_k = k^{-r}$, $\quad r > 0, \quad k = 1,2,\dots.$ Then*

$$N_{s+1} = [2^{1/r}N_s] + 1 \quad \textit{and} \quad n_s = [2^{1/r}N_s] + 1 - N_s.$$

*What implies*

$$N_s \asymp 2^{s/r} \quad \textit{and} \quad n_s \asymp 2^{s/r}.$$

**Example 3.3.** *Fix $0 < b < 1$ and take $\epsilon_k = 2^{-k^b}$, $\quad k = 0,1,2,\dots.$ Then*

$$N_s = s^{1/b} + O(1) \quad \textit{and} \quad n_s \asymp s^{1/b-1}.$$

Let $f \in L_p$. Rearrange the sequence $\|f_n\psi_n\|_p$ in decreasing order

$$\|f_{n_1}\psi_{n_1}\|_p \ge \|f_{n_2}\psi_{n_2}\|_p \ge \dots$$

and denote

$$a_k(f,p) := \|f_{n_k}\psi_{n_k}\|_p.$$

We give now some inequalities for $a_k(f,p)$ and $\sigma_m(f,\Psi)_p$. We will use brief notation $\sigma_m(f)_p := \sigma_m(f,\Psi)_p$ and $\sigma_0(f)_p := \|f\|_p$.

**Lemma 3.2.** *For any two positive integers $N < M$ we have*

$$a_M(f,p) \le C(p,\Psi)\sigma_N(f)_p(M-N)^{-1/p}.$$

**Lemma 3.3.** *For any sequence $m_0 < m_1 < m_2 < \dots$ of nonnegative integers we have*

$$\sigma_{m_s}(f)_p \le C(p,\Psi)\sum_{l=s}^{\infty} a_{m_l}(f,p)(m_{l+1}-m_l)^{1/p}.$$

**Theorem 3.5.** *Assume a given sequence $\mathcal{E} \in MDP$ satisfies the conditions*

$$\epsilon_{N_s} \geq C_1 2^{-s}, \qquad n_{s+1} \leq C_2 n_s, \qquad s = 0, 1, 2, \dots.$$

*Then we have the equivalence*

$$\sigma_n(f)_p \ll \epsilon_n \quad \iff \quad a_{N_s}(f,p) \ll 2^{-s} n_s^{-1/p}.$$

**Corollary 3.1.** *Theorem 3.5 applied to Examples 3.2, 3.3 gives the following relations:*

$$\sigma_m(f)_p \ll (m+1)^{-r} \quad \iff \quad a_n(f,p) \ll n^{-r-1/p}, \tag{3.6}$$

$$\sigma_m(f)_p \ll 2^{-m^b} \quad \iff \quad a_n(f,p) \ll 2^{-n^b} n^{(1-1/b)/p}. \tag{3.7}$$

**Remark 3.3.** *Making use of Lemmas 3.2 and 3.3 we can prove a version of Corollary 3.1 with the sign $\ll$ replaced by $\asymp$.*

Theorem 3.5 and Corollary 3.1 are in spirit of classical Jackson-Bernstein direct and inverse theorems in linear approximation theory, where conditions on the corresponding sequences of approximating characteristics are imposed in the form

$$E_n(f)_p \ll \epsilon_n, \quad \text{or} \quad \|E_n(f)_p/\epsilon_n\|_{l_\infty} < \infty. \tag{3.8}$$

It is well known (see [D]) that in studying many questions of approximation theory it is convenient to consider along with restriction (3.8) the following its generalization

$$\|E_n(f)_p/\epsilon_n\|_{l_q} < \infty. \tag{3.9}$$

Lemmas 3.2 and 3.3 are also useful in considering this more general case. For instance, in the particular case of Example 3.2 one gets the following statement.

**Theorem 3.6.** *Let $1 < p < \infty$ and $0 < q < \infty$. Then for any positive $r$ we have the equivalence relation*

$$\sum_m \sigma_m(f)_p^q m^{rq-1} < \infty \quad \iff \quad \sum_n a_n(f,p)^q n^{rq-1+q/p} < \infty.$$

**Remark 3.4.** *The condition*

$$\sum_n a_n(f,p)^q n^{rq-1+q/p} < \infty$$

*with $q = \beta := (r+1/p)^{-1}$ takes a very simple form*

$$\sum_n a_n(f,p)^\beta = \sum_n \|f_n \psi_n\|_p^\beta < \infty. \tag{3.10}$$

*In the case $\Psi = \mathcal{H}_p$ the condition (3.10) is equivalent to $f$ is in Besov space $B^r_\beta(L_\beta)$.*

**Corollary 3.2.** *Theorem 3.6 implies the following relation*

$$\sum_m \sigma_m(f,\mathcal{H})_p^\beta m^{r\beta-1} < \infty \qquad \Longleftrightarrow \qquad f \in B^r_\beta(L_\beta),$$

*where* $\beta := (r+1/p)^{-1}$.

The statement similar to Corollary 3.2 for free knots spline approximation was proved by P. Petrushev [P]. Corollary 3.2 and further results in this direction can be found in [DP] and [DJP]. We want to remark here that conditions in terms of $a_n(f,p)$ are convenient in applications. For instance, the relation (3.6) can be rewritten using the idea of thresholding. For a given $f \in L_p$ denote

$$T(\epsilon) := \#\{a_k(f,p) \quad : \quad a_k(f,p) \geq \epsilon\}.$$

Then (3.6) is equivalent to

$$\sigma_m(f)_p \ll (m+1)^{-r} \qquad \Longleftrightarrow \qquad T(\epsilon) \ll \epsilon^{-(r+1/p)^{-1}}.$$

For further results in this direction see [D], [CDH], [Os].

**3.4 Stability.** In this section we assume that a basis $\Psi = \{\psi_k\}_{k=1}^\infty$ is an unconditional normalized ($\|\psi_k\| = 1, k = 1,2,\dots$) basis for $X$ (see Definition 3.2).

The uniform boundedness principle implies that the unconditional constant

$$K := K(X,\Psi) := \sup_\theta \|M_\theta\|$$

is finite.

The following theorem is a well known fact about unconditional bases (see [LT], p.19).

**Theorem 3.7.** *Let $\Psi$ be an unconditional basis for $X$. Then for every choice of bounded scalars $\{\lambda_k\}_{k=1}^\infty$, we have*

$$\|\sum_{k=1}^\infty \lambda_k a_k \psi_k\| \leq 2K \sup_k |\lambda_k| \|\sum_{k=1}^\infty a_k \psi_k\|$$

*(in the case of real Banach space $X$ we can take $K$ instead of $2K$).*

In numerical implementation of nonlinear $m$-term approximation one usually prefers to employ the strategy known as thresholding (see [D, S.7.8]) instead of greedy algorithm. We define and study here the soft thresholding. Let a real function $v(x)$ defined for $x \geq 0$ satisfies the following relations

$$v(x) = \begin{cases} 1, & \text{for } x \geq 1 \\ 0, & \text{for } 0 \leq x \leq 1/2, \end{cases} \tag{3.11}$$

$$|v(x)| \leq A, \quad x \in [0,1]; \tag{3.12}$$

there is a constant $C_L$ such that for any $x,y\in[0,\infty)$ we have

$$|v(x)-v(y)|\le C_L|x-y|. \tag{3.13}$$

Let

$$f=\sum_{k=1}^{\infty}c_k(f)\psi_k.$$

We define a soft thresholding mapping $T_{\epsilon,v}$ as follows. Take $\epsilon>0$ and set

$$T_{\epsilon,v}(f):=\sum_k v(|c_k(f)|/\epsilon)c_k(f)\psi_k.$$

Theorem 3.7 implies that

$$\|T_{\epsilon,v}(f)\|\le 2KA\|f\|. \tag{3.14}$$

It was proved in [T21] that the mapping $T_{\epsilon,v}$ satisfies the Lipschitz condition with a constant independent of $\epsilon$.

**Theorem 3.8.** *For any $\epsilon$ and any functions $f,g\in X$ we have*

$$\|T_{\epsilon,v}(f)-T_{\epsilon,v}(g)\|\le(3A+2C_L)2K\|f-g\|.$$

**Open problems.**

**3.1.** Does the inequality

$$\|f-G_m^{c,t}(f,\mathcal{RT})\|_p\le C_1(p,t)\sigma_n(f,\mathcal{RT})_p$$

hold for any $f\in L_p(\mathbb{T})$, $1<p<\infty$, with $m\le C_2(p,t)n$?

**3.2.** Does the inequality

$$\|f-G_m^{c,t}(f,\mathcal{H}_p)\|_p\le C_1(p,t)\sigma_n(f,\mathcal{H}_p)_p$$

hold for any $f\in L_p(0,1)$, $1<p<\infty$, with $m\le C_2(p,t)n$?

**3.3.** Find the order of the quantity

$$\sup_{f\in W_p^r}\|f-G_m^{c,t}(f,\mathcal{RT})\|_p,\quad 1<p<\infty.$$

**3.4.** Find greedy type algorithm realizing near best approximation in the $L_p([0,1]^d)$, $1<p<\infty$, $d\ge2$, with regard to $\mathcal{H}_p^d$ for individual functions.

## 4. Some Convergence Results

In Section 3 we discussed greedy bases. That is justified from the point of view of efficient approximation. It follows from Proposition 3.1 that the inequality

$$\|G_m(f,\Psi,\rho)\|\le(G+1)\|f\| \tag{4.1}$$

holds for all $m$ and all $f\in X$ for every $\rho\in D(f)$.

**Definition 4.1.** *We say that a basis $\Psi$ is quasi-greedy if there exists a constant $C_Q$ such that for any $f \in X$ and any finite set of indices $\Lambda$, having the property*

$$\min_{k\in\Lambda}|c_k(f)| \geq \max_{k\notin\Lambda}|c_k(f)|, \tag{4.2}$$

*we have*

$$\|S_\Lambda(f,\Psi)\| = \|\sum_{k\in\Lambda} c_k(f)\psi_k\| \leq C_Q\|f\|. \tag{4.3}$$

It is clear that the inequalities (4.1) and (4.3) are equivalent. P. Wojtaszczyk [W2] proved that a basis $\Psi$ is quasi-greedy if and only if the sequence $\{G_m(f,\Psi,\rho)\}$ converges to $f$ for all $f \in X$ and any $\rho \in D(f)$. We constructed in [KoT1] an example of quasi-greedy basis that is not an unconditional basis (and, therefore, not a greedy basis). We have the following theorem for the trigonometric system.

**Theorem 4.1.** *The trigonometric system $\mathcal{T}$ is not a quasi-greedy basis for $L_p$ if $p \neq 2$.*

This theorem has been proved in [T19] and for $p < 2$ it has been proved independently and by different method in [CF]. We mention here that the method from [T19] gives a little more than stated in Theorem 4.1.

**Theorem 4.2.** *There exists a continuous function $f$ such that $G_m(f,\mathcal{T})$ does not converge to $f$ in $L_p$ for any $p > 2$.*

**Theorem 4.3.** *There exists a function $f$ that belongs to any $L_p$, $p < 2$, such that $G_m(f,\mathcal{T})$ does not converge to $f$ in measure.*

The proof of both theorems is based on two examples (one for $p > 2$ and the other for $p < 2$) constructed in [T19, pp 574 575]. We prove here only Theorem 4.3 where we use the example from [T19] for $p < 2$.

*Proof of Theorem 4.3.* We use the Rudin-Shapiro polynomials (see [KS])

$$R_N(x) = \sum_{k=0}^{N-1}\epsilon_k e^{ikx}, \quad \epsilon_k = \pm 1, \quad x \in \mathbb{T},$$

that satisfy the inequality

$$\|R_N\|_\infty \leq CN^{1/2}, \tag{4.4}$$

with an absolute constant $C$. Denote for $s = \pm 1$

$$\Lambda_s(N) := \{k \quad : \quad \hat{R}_N(k) = s\}.$$

Denote also

$$D_\Lambda(x) := \sum_{k\in\Lambda} e^{ikx}.$$

Then

$$R_N = D_{\Lambda_{+1}} - D_{\Lambda_{-1}}.$$

The inequality (4.4) implies

$$\|R_N\|_1 \geq C_1 N^{1/2}.$$

Using this inequality we prove that there exist two positive constants $c_1$ and $c_2$ such that for one of $s = \pm 1$ we have

$$m\{x \quad : \quad |D_{\Lambda_s(N)}(x)| \geq c_1 N^{1/2}\} \geq c_2. \tag{4.5}$$

We define a function $f$ from Theorem 4.3 as follows

$$f := \sum_{v=1}^{\infty} 2^{-v/2} e^{i2^v x}(D_{\Lambda_s(2^v)} + s2^{-v} R_{2^v}).$$

Then for appropriately chosen $m_1$ and $m_2$ we get

$$G_{m_1}(f, \mathcal{T}) - G_{m_2}(f, \mathcal{T}) = 2^{-v/2} e^{i2^v x}(1 + 2^{-v}) D_{\Lambda_s(2^v)}$$

and by (4.5)

$$m\{x \quad : \quad |G_{m_1}(f) - G_{m_2}(f)| \geq c_1\} \geq c_2$$

what shows that $\{G_m(f, \mathcal{T})\}$ does not converge in measure. Further, for any $1 < p < 2$ we have

$$\|D_{\Lambda_s(2^v)} + s2^{-v} R_{2^v}\|_p \leq C 2^{v(1-1/p)}$$

what implies that $f \in L_p$.

We also mention two interesting results on convergence almost everywhere. T.W. Körner answering a question raised by Carleson and Coifman constructed in [K1] a function from $L_2$ and then in [K2] a continuous function such that $\{G_m(f, \mathcal{T})\}$ diverges almost everywhere. T. Tao [Ta] proved that for the Haar system we have convergence: the sequence $\{G_m(f, \mathcal{H}_p)\}$ converges almost everywhere to $f$ for any $f \in L_p$, $1 < p < \infty$.

**Open problems.**

**4.1.** Does $L_p$-Greedy Algorithm with regard to $\mathcal{T}$ converge in $L_p$, $1 < p < \infty$, for each $f \in L_p(\mathbb{T})$?

**4.2.** Does Dual Greedy Algorithm with regard to $\mathcal{T}$ converge in $L_p$, $1 < p < \infty$, for each $f \in L_p(\mathbb{T})$?

**4.3.** Does $L_p$-Greedy Algorithm with regard to $\mathcal{H}_p$ converge in $L_p$, $1 < p < \infty$, for each $f \in L_p(0,1)$?

**4.4.** Does Dual Greedy Algorithm with regard to $\mathcal{H}_p$ converge in $L_p$, $1 < p < \infty$, for each $f \in L_p(0,1)$?

## 5. Widths. Optimal methods in Linear Approximation

In Sections 2 and 3 a basis $\Psi$ was chosen a priori. In many problems when an application to physical or engineering problems dictates the choice of a basis it is the case. However, in many other problems we can choose an appropriate basis for approximation. This leads to a search for optimal bases of approximation. The

first result in this direction was obtained by Kolmogorov. In 1936 Kolmogorov introduced the concept of width of a class $F$ in a space $X$ :

$$d_n(F,X) = \inf_{\{\phi_j\}_{j=1}^n} \sup_{f\in F} \inf_{\{c_j\}_{j=1}^n} \|f - \sum_{j=1}^n c_j\phi_j\|_X.$$

This concept allows us to find for fixed $n$ and for a class $F$ a subspace of dimension $n$, optimal with respect to the construction of an approximating element as the element of best approximation. In other words, the concept of width allows us to choose among various Chebyshev methods having the same quantitative characteristic of complexity (the dimension of the approximating subspace) the one which has the greatest accuracy. The first result about widths, namely Kolmogorov's result (1936)

$$d_{2n+1}(W_2^r, L_2) = (n+1)^{-r},$$

showed that the best subspace of dimension $2n+1$ for approximation of classes of periodic functions is the subspace of trigonometric polynomials of order $n$. This result confirmed that the approximation of functions in the class $W_2^r$ by trigonometric polynomials is natural. Further estimates of the widths $d_{2n+1}(W_q^r, L_p)$, $1 \le q$, $p \le \infty$, some of which are discussed here, showed that for some values of the parameters $q$, $p$ the subspace of trigonometric polynomials of order $n$ is optimal (in the sense of order) but for other values of $q,p$ this subspace is not optimal.

The Ismagilov [I] estimate for the quantity $d_n(W_1^r, L_\infty)$ gave the first example where the subspace of trigonometric polynomials of order $n$ is not optimal. This phenomenon was thoroughly studied by Kashin [Ka1].

In analogy to the concept of Kolmogorov width, that is, to the problem concerning the best Chebyshev method, the problems concerning the best linear method and the best Fourier method were considered.

Tikhomirov [Ti] introduced the concept of linear width :

$$\lambda_n(F,X) = \inf_{A:\text{rank}\, A\le n} \sup_{f\in F} \|f - Af\|_X,$$

and the concept of orthowidth (Fourier width) was introduced in [T4]

$$\varphi_n(F,X) := d_n^\perp(F,X) := \inf_{\text{orthonormal system } \{u_i\}_{i=1}^n} \sup_{f\in F} \left\| f - \sum_{i=1}^n \langle f, u_i\rangle u_i \right\|_X .$$

All these widths have as a starting point a function class $F$. Thus in this setting we choose a priori a function class $F$ and look for optimal subspaces for approximation of a given class. The following results are well known [Te2]. We present these results for $r$ positive integer. Similar results hold for any $r$ greater than some $\alpha(q,p) \le 1$, which is defined below in Theorem 5.1. Positive integers satisfy the inequality $r > \alpha(q,p)$ for all $1 \le q,p \le \infty$, except $q=1$, $p=\infty$ where we have $\alpha(1,\infty) = 1$. Thus in the case $q=1$, $p=\infty$ we assume $r > 1$.

**A.** In the case $1 \le p \le q \le \infty$ or $1 \le q \le p \le 2$ one has

$$\varphi_n(W_q^r, L_p) \asymp \lambda_n(W_q^r, L_p) \asymp d_n(W_q^r, L_p) \asymp n^{-r+(1/q-1/p)_+}. \tag{5.1}$$

**B.** In the case $1 \le q < p \le \infty$, $p > 2$, one has

$$d_n(W^r_q, L_p) \asymp n^{-r+(1/q-1/2)_+},$$

$$\lambda_n(W^r_q, L_p) \asymp n^{-r+\max(1/q-1/2,1/2-1/p)},$$

$$\varphi_n(W^r_q, L_p) \asymp n^{-r+1/q-1/p}.$$

In the case A the classical trigonometric system provides the optimal orders for all widths, except $\varphi_n$ for $q = p = 1, \infty$. Let us discuss a more interesting case B for a particular choice of $q = 2$ and $p = \infty$. We have

$$d_n(W^r_2, L_\infty) \asymp n^{-r}, \tag{5.2}$$

$$\lambda_n(W^r_2, L_\infty) \asymp \varphi_n(W^r_2, L_\infty) \asymp n^{-r+1/2}. \tag{5.3}$$

These relations show that if we drop the linearity requiment for approximation method we gain in accuracy a factor $n^{-1/2}$. However, there is a big difficulty in realization of the estimate (5.2). We know by Kashin's result that there exists a subspace realizing (5.2) but we do not know a way to construct it. Thus it is only an existence theorem for now.

Let us discuss one more special case $q = 1$ and $p = \infty$. In this case we have

$$d_n(W^r_1, L_\infty) \asymp \lambda_n(W^r_1, L_\infty) \asymp n^{-r+1/2} \tag{5.4}$$

and

$$\varphi_n(W^r_1, L_\infty) \asymp n^{-r+1}. \tag{5.5}$$

Therefore, by (5.4) the best possible approximation (in the sense of order) can be realized by linear method, say, $A_n$. However, by (5.5) this linear method $A_n$ is certainly not an orthogonal projector. Moreover, by [Te2] it can not satisfy even the following much weaker restriction $\|A_n(e^{ikx})\|_2 \le C$, $k \in \mathbb{Z}$. This means that the optimal linear operator $A_n$ is unstable. A small change in some of Fourier coefficients of $f$ may result in a big change of $\|A_n(f)\|_2$.

Let us make some conclusions now. In Linear Approximation of $W^r_q$ in $L_p$ the bottom line is given by $\varphi_n(W^r_q, L_p)$ where the approximation method is the simplest orthogonal projection. Partial sums with regard to classical systems provide an optimal error of approximation for this width. The trigonometric system works for all $1 \le q, p \le \infty$ except $(q, p) = (1, 1), (\infty, \infty)$. The wavelet systems (see [AT]) work for all $1 \le q, p \le \infty$. On the example of the pair $(W^r_1, L_\infty)$ we have seen that we need to sacrifice important and convenient properties of approximating operator in order to achieve better accuracy. On the example of $(W^r_2, L_\infty)$ we have seen that we need to pay even bigger price for better accuracy in a form of proving only an existence theorem instead of providing a constructive method of approximation.

Let us continue the discussion from Section 2 on interplay between approximation of individual functions and function classes. Let us first try to associate with an individual function $f$ a sequence of the Kolmogorov widths. It is clear that the choice $F[f] := \{f\}$ does not work because $d_1(F[f]) = 0$ for each $f$. The idea is to

find a minimal reasonable class that contains $f$. In the periodic case it is natural to associate with $f(x)$ all translates $f(x-y)$. Thus define $F[f] := \{f(x-y), \quad y \in \mathbb{T}\}$. All known classes of periodic functions are shift invariant. In such a case we have for $f \in F$ that $F[f] \subset F$ and $d_n(F[f], X) \le d_n(F, X)$. We will present some results from [T5]. For $r \in \mathbb{Z}_+$, $\alpha \in \mathbb{R}_+$ denote

$$W^r H^\alpha_q := \{f \quad : \quad f^{(r)} - \text{absolutely continuous},$$

$$\|f^{(r)}(x) - f^{(r)}(y)\|_q \le |x-y|^\alpha, \quad x, y \in \mathbb{T}\}.$$

**Theorem 5.1.** *Let $1 \le q \le p \le \infty$ or $2 \le p \le q \le \infty$. Then each class $W^r H^\alpha_q$ with $0 < \alpha < 1$ and $r + \alpha \ge \alpha(q,p)$ contains a function $f$ such that*

$$\liminf_{n\to\infty} d_n(F[f], L_p)/d_n(W^r H^\alpha_q, L_p) > 0.$$

*We define here $\alpha(q,p) := (1/q - 1/p)_+$ for $1 \le q \le p \le 2$, $2 \le p \le q \le \infty$ and $\alpha(q,p) := \max(1/q, 1/2)$ for $1 \le q \le p \le \infty$, $p > 2$.*

Let us consider one particular case $q = p = \infty$, $\alpha = 1$, that is not covered by Theorem 5.1. As established by Tikhomirov [Ti], the values of the Kolmogorov width in this case are given by approximations by trigonometric polynomials. Results of Nikol'skii and the author mentioned in Section 2 show that each class $W^r_\infty$ contains a function asymptotically extreme for the best approximation by trigonometric polynomials. It turns that the picture is different for the asymptotic behavior of the widths $d_n(F[f], L_\infty)$.

**Theorem 5.2.** *Any function $f \in W^r_\infty$, $r > 1/2$ satisfies*

$$d_n(F[f], L_\infty) = o(d_n(W^r_\infty, L_\infty)).$$

It is intersting to note that for any periodic function $f \in L_p(\mathbb{T})$ we have

$$\sigma_m(f(x-y), \Pi)_{p,\infty} = d_m(F[f], L_p) \le \sigma_m(f, \mathcal{T})_p. \tag{5.6}$$

It is proved in [T5] that for $1 \le q \le p \le \infty$ one has

$$d^{ind}_m(W^r H^\alpha_q, L_p) := \sup_{f\in W^r H^\alpha_q} d_m(F[f], L_p) \asymp \tag{5.7}$$

$$d_m(W^r H^\alpha_q, L_p) \asymp m^{-r-\alpha+(1/q-\max(1/2,1/p))_+}$$

provided $r + \alpha > \alpha(q,p)$ with $\alpha(q,p)$ defined in Theorem 5.1. We proved in [DT1] that

$$\sigma_m(W^r H^\alpha_q, \mathcal{T})_p \asymp m^{-r-\alpha+(1/q-\max(1/2,1/p))_+} \tag{5.8}$$

under the same assumption $r + \alpha > \alpha(q,p)$. Relations (5.7) and (5.8) show that for any pair of $(q,p)$, $1 \le q \le p \le \infty$, and for each function $f \in W^r H^\alpha_q$ the trigonometric system $\mathcal{T}$ provides a subspace $\mathcal{T}(\Lambda)$, $\#\Lambda \le m$, with frequences in $\Lambda$ such that

$$d_m(F[f], L_p) \le \sup_{y\in\mathbb{T}} \inf_{t\in\mathcal{T}(\Lambda)} \|f(\cdot - y) - t(\cdot)\|_p \ll d^{ind}_m(W^r H^\alpha_q, L_p).$$

**Open problems.**

**5.1.** Construct a subspace realizing (5.2).

**5.2.** Does there exist $f \in W^r H^\alpha_\infty$, $0 < \alpha < 1$, such that

$$d_n(F[f], L_1) \gg n^{-r-\alpha}?$$

## 6. Optimal Methods in Nonlinear Approximation

In the widths problem of Linear Approximation we were looking for an optimal $n$-dimensional subspace for approximating a given function class. A nonlinear analog of this setting is the following. Let a function class $F$ and a Banach space $X$ be given. Assume that on the base of some additional information we know that our basis for $m$-term approximation should satisfy some structural properties, for instance, has to be orthogonal. Then similarly to the setting for the widths $d_n$, $\lambda_n$, $\varphi_n$ we get the optimization problems for $m$-term nonlinear approximation (see Introduction). Let $\mathbb{B}$ be a collection of bases satisfying a given property.

**I.** Define an analog of the Kolmogorov width

$$\sigma_m(F,\mathbb{B})_X := \inf_{\Psi\in\mathbb{B}} \sup_{f\in F} \sigma_m(f,\Psi)_X.$$

**II.** Define an analog of the orthowidth

$$\gamma_m(F,\mathbb{B})_X := \inf_{\Psi\in\mathbb{B}} \sup_{f\in F} \|f-G_m(f,\Psi)\|_X.$$

We present here some results in the case $\mathbb{B}=\mathbb{O}$ - the set of orthonormal bases, $F=W^r_q$, $X=L_p$, $1\le q,p\le\infty$. First of all we formulate a result (see [KT1], [T18]) that shows that in the case $p<2$ we need some more restrictions on $\mathbb{B}$ in order to get meaningful results (lower bounds).

**Proposition 6.1.** *For any $1\le p<2$ there exists a complete in $L_2(0,1)$ orthonormal system $\Phi$ such that for each $f\in L_p(0,1)$ we have $\sigma_1(f,\Phi)_p=0$.*

Let us restrict our further discussion to the case $p\ge 2$. This case was also more interesting in the Linear Approximation discussion (see Section 5). Kashin [Ka2] proved that

$$\sigma_m(W^r_\infty,\mathbb{O})_2 \gg m^{-r}. \tag{6.1}$$

We proved (see [DT1]) that

$$\sigma_m(W^r_2,\mathcal{T})_\infty \ll m^{-r}. \tag{6.2}$$

The estimates (6.1) and (6.2) imply that for $2\le q,p\le\infty$ we have

$$\sigma_m(W^r_q,\mathbb{O})_p \asymp \sigma_m(W^r_q,\mathcal{T})_p \asymp m^{-r}. \tag{6.3}$$

Let us compare this relation with (5.2). We see that best $m$-term trigonometric approximation provides the same accuracy as the best approximation from an optimal $m$-dimensional subspace. An advantage of nonlinear approximation here is that we use a natural basis instead of existing but nonconstructive subspace. However, we should note that the estimate (6.2) was proved in [DT1] as an existence theorem. We did not give an algorithm to get (6.2) in [DT1] and do not know it now. The Thresholding Greedy Algorithm does not provide the estimate (6.2). We have (see [T19])

$$\sup_{f\in W^r_2} \|f-G_m(f,\mathcal{T})\|_\infty \asymp m^{-r+1/2}.$$

It is known from different results (see [DJP], [D], [T21]) that wavelets are well designed for nonlinear approximation. We present here one general result in this direction. We consider a basis $\Psi := \{\psi_I\}_{I\in D}$ enumerated by dyadic intervals $I$ of $[0,1]^d$, $I = I_1 \times \cdots \times I_d$, $I_j$ is a dyadic interval of $[0,1]$, $j = 1,\dots,d$, which satisfies certain properties. Let $L_p := L_p(\Omega)$ with normalized Lebesgue measure on $\Omega$, $|\Omega| = 1$. First of all we assume that for all $1 < q,p < \infty$ and $I \in D$, $D := D([0,1]^d)$ is the set of all dyadic intervals of $[0,1]^d$, we have

$$\|\psi_I\|_p \asymp \|\psi_I\|_q |I|^{1/p-1/q}, \tag{6.4}$$

with constants independent of $I$. This property can be easily checked for a given basis.

Next, assume that for any $s = (s_1,\dots,s_d) \in \mathbb{Z}^d$, $s_j \ge 0$, $j = 1,\dots,d$, and any $\{c_I\}$ we have for $1 < p < \infty$

$$\|\sum_{I\in D_s} c_I\psi_I\|_p^p \asymp \sum_{I\in D_s} \|c_I\psi_I\|_p^p, \tag{6.5}$$

where

$$D_s := \{I = I_1\times\cdots\times I_d \in D \quad : \quad |I_j| = 2^{-s_j}, \quad j = 1,\dots,d\}.$$

This assumption allows us to estimate the $L_p$-norm of a dyadic block in terms of Fourier coefficients.

The third assumption is that $\Psi$ is a basis satisfying the Littlewood-Paley inequality. This means the following. Let $1 < p < \infty$ and $f \in L_p$ has an expansion

$$f = \sum_I f_I\psi_I.$$

We assume that

$$\lim_{\min_j \mu_j\to\infty} \|f - \sum_{s_j\le\mu_j, j=1,\dots,d}\ \sum_{I\in D_s} f_I\psi_I\|_p = 0, \tag{6.6}$$

and

$$\|f\|_p \asymp \|(\sum_s |\sum_{I\in D_s} f_I\psi_I|^2)^{1/2}\|_p. \tag{6.7}$$

Let $\mu \in \mathbb{Z}^d$, $\mu_j \ge 0$, $j = 1,\dots,d$. Denote by $\Psi(\mu)$ the subspace of polynomials of the form

$$\psi = \sum_{s_j\le\mu_j, j=1,\dots,d}\ \sum_{I\in D_s} c_I\psi_I.$$

We define now a function class. Let $R = (R_1,\dots,R_d)$, $R_j > 0$, $j = 1,\dots,d$, and

$$g(R) := (\sum_{j=1}^d R_j^{-1})^{-1}.$$

For natural numbers $l$ denote

$$\Psi(R,l) := \Psi(\mu), \qquad \mu_j = [g(R)l/R_j], \quad j = 1,\dots,d.$$

We define the class $H_q^R(\Psi)$ as the set of functions $f \in L_q$ representable in the form

$$f = \sum_{l=1}^\infty t_l, \quad t_l \in \Psi(R,l), \quad \|t_l\|_q \le 2^{-g(R)l}.$$

**Theorem 6.1.** *Let $1 < q, p < \infty$ and $g(R) > (1/q - 1/p)_+$. Then for $\Psi$ satisfying (6.4) (6.7) we have*

$$\sup_{f\in H^R_q(\Psi)} \|f - G^{L_p}_m(f,\Psi)\|_p \ll m^{-g(R)}.$$

In the periodic case the following basis $U^d := U\times\cdots\times U$ can be taken in place of $\Psi$ in Theorem 6.1. We define the system $U := \{U_I\}$ in the univariate case. Denote

$$U^+_n(x) := \sum_{k=0}^{2^n-1} e^{ikx} = \frac{e^{i2^nx}-1}{e^{ix}-1}, \quad n = 0,1,2,\dots;$$

$$U^+_{n,k}(x) := e^{i2^nx}U^+_n(x - 2\pi k2^{-n}), \quad k = 0,1,\dots,2^n-1;$$

$$U^-_{n,k}(x) := e^{-i2^nx}U^+_n(-x + 2\pi k2^{-n}), \quad k = 0,1,\dots,2^n-1.$$

We normalize the system of functions $\{U^+_{n,k}, U^-_{n,k}\}$ in $L_2$ and enumerate it by dyadic intervals. We write

$$U_I(x) := 2^{-n/2}U^+_{n,k}(x) \quad \text{with} \quad I = [(k+1/2)2^{-n}, (k+1)2^{-n});$$

$$U_I(x) := 2^{-n/2}U^-_{n,k}(x) \quad \text{with} \quad I = [k2^{-n}, (k+1/2)2^{-n});$$

and

$$U_{[0,1)}(x) := 1.$$

It is well known that $H^R_q(U^d)$ is equivalent to the standard anisotropic multivariate periodic Hölder-Nikol'skii classes $NH^R_p$. We define these classes in the following way. The class $NH^R_p$, $R = (R_1,\dots,R_d)$ and $1 \le p \le \infty$, is the set of periodic functions $f \in L_p([0,2\pi]^d)$ such that for each $l_j = [R_j]+1$, $j = 1,\dots,d$, the following relations hold

$$\|f\|_p \le 1, \qquad \|\Delta^{l_j,j}_t f\|_p \le |t|^{R_j}, \quad j = 1,\dots,d, \tag{6.8}$$

where $\Delta^{l,j}_t$ is the $l$-th difference with step $t$ in the variable $x_j$. In the case $d=1$ $NH^R_p$ coincides with the standard Hölder class $H^R_p$. Then Theorem 6.1 gives.

**Theorem 6.2.** *Let $1 < q, p < \infty$; then for $R$ such that $g(R) > (1/q - 1/p)_+$ we have*

$$\sup_{f\in NH^R_q} \|f - G^{L_p}_m(f,U^d)\|_p \ll m^{-g(R)}.$$

We also proved in [T21] that the bais $U^d$ is an optimal orthonormal basis for approximation of classes $NH^R_q$ in $L_p$:

$$\sigma_m(NH^R_q,\mathbb{O})_p \asymp \sigma_m(NH^R_q,U^d)_p \asymp m^{-g(R)} \tag{6.9}$$

for $1 < q < \infty$, $2 \le p < \infty$, $g(R) > (1/q - 1/p)_+$. It is important to remark that Theorem 6.2 guaranties that the estimate in (6.9) can be realized by TGA with regard to $U^d$.

**Open problem 6.1.** Find a constructive proof of (6.2) (provide an algorithm).

## 7. Universality

In this section we discuss in the model case of anisotropic function classes a general approach formulated in Introduction of how to choose a good basis (dictionary) for approximation. This approach consists of several steps. We concentrate here on nonlinear approximation and compare realizations of this approach for linear and nonlinear approximations. The first step in this approach is an optimization problem. In both cases (linear and nonlinear) we begin with a function class $F$ in a given Banach space $X$. A classical example of optimization problem in the linear case is the problem of finding (estimating) the Kolmogorov width $d_m(F,X)$. This concept allows us to choose among various Chebyshev methods (best approximation) having the same dimension of the approximating subspace the one which has the best accuracy. The asymptotic behavior (in the sense of order) of the sequence $\{d_m(F,X)\}_{m=1}^{\infty}$ is known for a number of function classes and Banach spaces. It turned out that in many cases, for instance, in the case $F = W_p^r$ is a standard Sobolev class and $X = L_p$, the optimal (in the sense of order) $m$-dimensional subspaces can be formed as subspaces spanned by $m$ elements from one orthogonal system. We describe this for the multivariate periodic Hölder-Nikol'skii classes $NH_p^R$. It is known (see for instance [Te2]) that

$$d_m(NH_p^R, L_p) \asymp m^{-g(R)}, \quad 1 \le p \le \infty. \tag{7.1}$$

It is also known that the subspaces of trigonometric polynomials $\mathcal{T}(R,l)$ with frequences $k$ satisfying the inequalities

$$|k_j| \le 2^{g(R)l/R_j}, \quad j = 1, \dots, d,$$

can be chosen to realize (7.1). In this case $l$ is set to be the largest satisfying $\dim \mathcal{T}(R,l) \le m$. We stress here that optimal (in the sense of order) subspaces $\mathcal{T}(R,l)$ are different for different $R$ and formed from the same (trigonometric) system.

A nonlinear analog of the Kolmogorov $m$-width setting was discussed in Section 6. In this section we consider only the case $\mathbb{D} = \mathbb{O}$ the set of all orthogonal bases on a given domain. In Section 6 we mentioned that

$$\sigma_m(NH_q^R, \mathbb{O})_{L_p} \asymp m^{-g(R)} \tag{7.2}$$

for

$$1 < q < \infty, \quad 2 \le p < \infty, \quad g(R) > (1/q - 1/p)_+.$$

It is important to remark that the basis $U^d$ realizes (7.2) for all $R$ (see the definition of $U^d$ in Section 3).

The second step in our approach is to look for a universal basis (dictionary) for approximation. The mentioned above result on the basis $U^d$ means that $U^d$ is universal for the pair $(\mathcal{F}_q([A,B]), \mathbb{O})$ and the space $X = L_p([0,2\pi]^d)$ for $A, B \in \mathbb{Z}_+^d$ such that $g(A) > (1/q - 1/p)_+$, $1 < q < \infty$, $2 \le p < \infty$, where

$$\mathcal{F}_q([A,B]) := \{NH_q^R \quad : \quad 0 < A_j \le R_j \le B_j < \infty, \quad j = 1, \dots, d\}.$$

It is interesting to compare this result on universal bases in nonlinear approximation with the corresponding result in the linear setting. We define the index $\kappa(m, \mathcal{F}, X)$ of universality for a collection $\mathcal{F}$ with respect to the Kolmogorov width in $X$:

$$\kappa(m, \mathcal{F}, X) := L(m, \mathcal{F}, X)/m,$$

where $L(m,\mathcal{F},X)$ is the smallest number among those $L$ for which there is a system of functions $\{\varphi_i\}_{i=1}^L$ such that for each $F\in\mathcal{F}$ we have

$$\sup_{f\in F}\inf_{c_1,\dots,c_L}\|f-\sum_{i=1}^L c_i\varphi_i\|\le d_m(F,X).$$

It is proved in [T8] (see also [Te2, Ch.3, S.5]) that for any $A,B\in\mathbb{Z}^d_+$ such that $B_j>A_j$, $j=1,\dots,d$, we have

$$\kappa(m,\mathcal{F}_p([A,B]),L_p)\gg(\log m)^{d-1},\quad 1<p<\infty. \tag{7.3}$$

The estimate (7.3) implies that there is no Chebyshev methods universal for a nontrivial collection of anisotropic function classes. Thus, from the point of view of existence of universal methods the nonlinear setting has an advantage over the linear setting.

After two steps of realizing our approach in the nonlinear approximation we get a universal dictionary $\mathcal{D}_u$ for a collection of function classes $\mathcal{F}$, say, $U^d$ for $\mathcal{F}_q([A,B])$. This means that the dictionary $\mathcal{D}_u$ is well desinged for best $m$-term approximation of functions from function classes in the given collection. The third step is to find an algorithm (theoretical first) to realize best (near best) $m$-term approximation with regard to $\mathcal{D}_u$. It turned out that in the model case of $\mathcal{F}_q([A,B])$ and the basis $U^d$ there is a simple algorithm which realizes near best $m$-term approximation for classes $NH^R_q$. This is Thresholding Greedy Algorithm (see Theorem 6.2).

Thus we have established that in the above model case the basis $U^d$ is optimal for nonlinear $m$-term approximation in a very strong sense. The following two features of $U^d$ are the most important: 1) $U^d$ is the tensor product of the univariate basis $U$; 2) the univariate basis $U$ is a wavelet type basis. It is known [W1] that $U$ is $L_p$-equivalent, $1<p<\infty$, to the Haar basis. Then by Theorem 3.1 $U$ is a greedy basis for $L_p$, $1<p<\infty$. The tensor product structure of $U^d$ is important in making $U^d$ a universal basis for a collection of anisotropic Hölder-Nikol'skii classes. It would be ideal if $U^d$ is a greedy basis for $L_p(\mathbb{T}^d)$, $1<p<\infty$. Unfortunately, it is not a case. We have that for $1<p<\infty$

$$\sup_{f\in L_p}\|f-G^p_m(f,U^d)\|_p/\sigma_m(f,U^d)_p\asymp(\log m)^{(d-1)|1/2-1/p|}. \tag{7.4}$$

This relation follows from its analog with $U^d$ replaced by the multivariate Haar system $\mathcal{H}^d:=\mathcal{H}\times\dots\times\mathcal{H}$. The lower estimate in (7.4) for $\mathcal{H}^d$ was proved by R. Hochmuth; the upper estimate in (7.4) for $\mathcal{H}^d$ was proved in the case $d=2$, $4/3\le p\le 4$, and was conjectured for all $d$, $1<p<\infty$, in [T15]. The conjecture was proved in [W2].

## 8. Greedy Algorithms in Hilbert spaces

Perhaps the first example of $m$-term approximation with regard to redundant dictionary was considered by E. Schmidt in 1907 [S] who considered the approximation of functions $f(x,y)$ of two variables by bilinear forms

$$\sum_{i=1}^m u_i(x)v_i(y)$$

in $L_2([0,1]^2)$. This problem is closely connected with properties of the integral operator

$$J_f(g) := \int_0^1 f(x,y)g(y)dy$$

with kernel $f(x,y)$. E. Schmidt [S] gave an expansion (known as the Schmidt expansion)

$$f(x,y) = \sum_{j=1}^{\infty} s_j(J_f)\phi_j(x)\psi_j(y) \tag{8.S}$$

where $\{s_j(J_f)\}$ is a nonincreasing sequence of singular numbers of $J_f$, i.e. $s_j(J_f) := \lambda_j(J_f^* J_f)^{1/2}$, $\{\lambda_j(A)\}$ is a sequence of eigenvalues of an operator $A$, $J_f^*$ is the adjoint operator to $J_f$. The two sequences $\{\phi_j(x)\}$ and $\{\psi_j(y)\}$ form orthonormal sequences of eigenfunctions of the operators $J_f J_f^*$ and $J_f^* J_f$ respectively. He also proved that

$$\|f(x,y) - \sum_{j=1}^{m} s_j(J_f)\phi_j(x)\psi_j(y)\|_{L_2} = \inf_{u_j, v_j \in L_2, \quad j=1,\dots,m} \|f(x,y) - \sum_{j=1}^{m} u_j(x)v_j(y)\|_{L_2}.$$

It follows from the Schmidt expansion that the above best bilinear approximation can be realized by the Pure Greedy Algorithm. This was observed and used in several papers (see [Po] for history).

Another problem of this type which is well known in statistics is the projection pursuit regression problem. We formulate the related results in the function theory language. The problem is to approximate in $L_2$ a given function $f \in L_2$ by a sum of ridge functions, i.e. by

$$\sum_{j=1}^{m} r_j(\omega_j \cdot x), \quad x, \omega_j \in \mathbb{R}^d, \quad j = 1, \dots, m,$$

where $r_j$, $j = 1, \dots, m$, are univariate functions. The following greedy type algorithm (projection pursuit) was proposed in [FS] to solve this problem. Assume functions $r_1, \dots, r_{m-1}$ and vectors $\omega_1, \dots, \omega_{m-1}$ have been determined after $m-1$ steps of algorithm. Choose at $m$-th step a unit vector $\omega_m$ and a function $r_m$ to minimize the error

$$\|f(x) - \sum_{j=1}^{m} r_j(\omega_j \cdot x)\|_{L_2}.$$

This is one more example of Pure Greedy Algorithm. The Pure Greedy Algorithm and some other versions of greedy type algorithms have been intensively studied recently (see [B], [DDGS], [DMA], [Du], [DT2], [DT3], [H], [J1], [J2], [T14 24]). In this section we discuss PGA and some its modifications which make them more ready for implementation. We call this new type of greedy algorithms Weak Greedy Algorithms (see Introduction for definitions of PGA and WGA).

If $H_0$ is a finite dimensional subspace of $H$, we let $P_{H_0}$ be the orthogonal projector from $H$ onto $H_0$. That is $P_{H_0}(f)$ is the best approximation to $f$ from $H_0$. We let $g(f) \in \mathcal{D}$ be an element from $\mathcal{D}$ which maximizes $|\langle f, g\rangle|$. We shall assume for simplicity that such a maximizer exists; if not suitable modifications are necessary (see Weak Orthogonal Greedy Algorithm below) in the algorithm that follows.

**Orthogonal Greedy Algorithm (OGA).** *We define* $R_0^o(f) := R_0^o(f, \mathcal{D}) := f$ *and* $G_0^o(f) := G_0^o(f, \mathcal{D}) := 0$. *Then for each* $m \ge 1$, *we inductively define*

$$\begin{aligned} H_m :=& H_m(f) := \operatorname{span}\{g(R_0^o(f)), \dots, g(R_{m-1}^o(f))\} \\ G_m^o(f) :=& G_m^o(f, \mathcal{D}) := P_{H_m}(f) \\ R_m^o(f) :=& R_m^o(f, \mathcal{D}) := f - G_m^o(f). \end{aligned}$$

We remark that for each $f$ we have

$$\|f - G_m^o(f, \mathcal{D})\| \le \|R_{m-1}^o(f) - G_1(R_{m-1}^o(f), \mathcal{D})\|. \tag{8.1}$$

Let a sequence $\tau = \{t_k\}_{k=1}^\infty$, $0 \le t_k \le 1$, be given. Following [T20] we define Weak Orthogonal Greedy Algorithm.

**Weak Orthogonal Greedy Algorithm (WOGA).** *We define* $f_0^{o,\tau} := f$. *Then for each* $m \ge 1$ *we inductively define:*

*1).* $\varphi_m^{o,\tau} \in \mathcal{D}$ *is any satisfying*

$$|\langle f_{m-1}^{o,\tau}, \varphi_m^{o,\tau}\rangle| \ge t_m \sup_{g \in \mathcal{D}} |\langle f_{m-1}^{o,\tau}, g\rangle|;$$

*2).*

$$G_m^{o,\tau}(f, \mathcal{D}) := P_{H_m^\tau}(f), \quad \textit{where} \quad H_m^\tau := \operatorname{Span}(\varphi_1^{o,\tau}, \dots, \varphi_m^{o,\tau});$$

*3).*

$$f_m^{o,\tau} := f - G_m^{o,\tau}(f, \mathcal{D}).$$

It is clear that $G_m^\tau$ and $G_m^{o,\tau}$ in the case $t_k = 1$, $k = 1, 2, \dots$, coincide with PGA $G_m$ and OGA $G_m^o$ respectively. It is also clear that WGA and WOGA are more ready for implementaion than PGA and OGA.

**8.1. Convergence.** The convergence of PGA and WGA with $t_k = t$, $0 < t < 1$, was established in [J1] and [RW]. The first sufficient condition on $\tau$ which includes sequences with $\liminf_{k\to\infty} t_k = 0$ was obtained in [T20].

**Theorem 8.1.** *Assume*

$$\sum_{k=1}^\infty \frac{t_k}{k} = \infty. \tag{8.2}$$

*Then for any dictionary* $\mathcal{D}$ *and any* $f \in H$ *we have*

$$\lim_{m\to\infty} \|f - G_m^\tau(f, \mathcal{D})\| = 0.$$

In [T20] we reduced the proof of convergence of WGA with weakness sequence $\tau$ to some properties of $l_2$-sequences with regard to $\tau$. Theorem 8.1 was derived from the following two statements proved in [T20].

**Proposition 8.1.** *Let* $\tau$ *be such that for any* $\{a_j\}_{j=1}^\infty \in l_2$, $a_j \ge 0$, $j = 1, 2, \dots$ *we have*

$$\liminf_{n\to\infty} a_n \sum_{j=1}^n a_j / t_n = 0.$$

*Then for any* $H$, $\mathcal{D}$, *and* $f \in H$ *we have*

$$\lim_{m\to\infty} \|f_m^\tau\| = 0.$$

**Proposition 8.2.** *If $\tau$ satisfies the condition (8.2) then $\tau$ satisfies the assumption of Proposition 8.1.*

The following simple necessary condition

$$\sum_{k=1}^{\infty} t_k^2 = \infty$$

was mentioned in [T20]. The first nontrivial necessary conditions were obtained in [LTe]. We proved in [LTe] the following theorem.

**Theorem 8.2.** *In the class of monotone sequences $\tau = \{t_k\}_{k=1}^{\infty}$, $1 \geq t_1 \geq t_2 \geq \dots \geq 0$, the condition (8.2) is necessary and sufficient for convergence of Weak Greedy Algorithm for each $f$ and all Hilbert spaces $H$ and dictionaries $\mathcal{D}$.*

The proof of this theorem is based on a special procedure which we called Equalizer. In [LTe] we gave an example of a class of sequences $\tau$ for which the condition (8.2) is not a necessary condition for convergence. We also proved in [LTe] a theorem which covers Theorem 8.1.

**Theorem 8.3.** *Assume*

$$\sum_{s=0}^{\infty} (2^{-s} \sum_{k=2^s}^{2^{s+1}-1} t_k^2)^{1/2} = \infty.$$

*Then for any dictionary $\mathcal{D}$ and any $f \in H$ we have*

$$\lim_{m \to \infty} \|f - G_m^{\tau}(f, \mathcal{D})\| = 0.$$

We proved in [T23] a criterion on $\tau$ for convergence of WGA. Let us introduce some notation.

We define by $\mathcal{V}$ the class of sequences $x = \{x_k\}_{k=1}^{\infty}$, $x_k \geq 0$, $k = 1, 2, \dots$, with the following property: there exists a sequence $0 = q_0 < q_1 < \dots$ such that

$$\sum_{s=1}^{\infty} \frac{2^s}{\Delta q_s} < \infty; \tag{8.3}$$

and

$$\sum_{s=1}^{\infty} 2^{-s} \sum_{k=1}^{q_s} x_k^2 < \infty, \tag{8.4}$$

where $\Delta q_s := q_s - q_{s-1}$.

**Theorem 8.4.** *The condition $\tau \notin \mathcal{V}$ is necessary and sufficient for convergence of Weak Greedy Algorithm with weakness sequence $\tau$ for each $f$ and all Hilbert spaces $H$ and dictionaries $\mathcal{D}$.*

The proof of the sufficient part of Theorem 8.4 is a refinement of the original proof of Jones [J1]. The study of the behavior of sequences $a_n \sum_{j=1}^{n} a_j$ for $\{a_j\}_{j=1}^{\infty} \in l_2$, $a_j \geq 0$, $j = 1, 2, \dots$, plays an important role in the proof. It turns out that the class $\mathcal{V}$ appears naturally in the study of the above mentioned sequences. We proved in [T23] the following theorem.

**Theorem 8.5.** *The following two conditions are equivalent*

$$\text{(C.1)} \qquad \tau \notin \mathcal{V},$$

$$\text{(C.2)} \qquad \forall \{a_j\}_{j=1}^{\infty} \in l_2, \quad a_j \geq 0, \quad \liminf_{n\to\infty} a_n \sum_{j=1}^{n} a_j/t_n = 0.$$

We give a result on convergence of WOGA now.

**Theorem 8.6.** *Assume*

$$\sum_{k=1}^{\infty} t_k^2 = \infty. \tag{8.5}$$

*Then for any dictionary $\mathcal{D}$ and any $f \in H$ we have*

$$\lim_{m\to\infty} \|f - G_m^{o,\tau}(f,\mathcal{D})\| = 0. \tag{8.6}$$

**Remark 8.1.** *It is easy to see that in the case $\mathcal{D} = \mathcal{B}$ - orthonormal basis the assumption (8.5) is also necessary for convergence (8.6) for all $f$.*

Theorems 8.4 and 8.6 show that conditions on the weakness sequence for convergence of WGA and WOGA are different.

**8.2 Rate of convergence.** For a general dictionary $\mathcal{D}$ we define the class of functions

$$\mathcal{A}_1^o(\mathcal{D}, M) := \{f \in H : f = \sum_{k\in\Lambda} c_k w_k, \quad w_k \in \mathcal{D}, \ \#\Lambda < \infty \text{ and } \sum_{k\in\Lambda} |c_k| \leq M\}$$

and we define $\mathcal{A}_1(\mathcal{D}, M)$ as the closure (in $H$) of $\mathcal{A}_1^o(\mathcal{D}, M)$. Furthermore, we define $\mathcal{A}_1(\mathcal{D})$ as the union of the classes $\mathcal{A}_1(\mathcal{D}, M)$ over all $M > 0$. For $f \in \mathcal{A}_1(\mathcal{D})$, we define the norm

$$|f|_{\mathcal{A}_1(\mathcal{D})}$$

as the smallest $M$ such that $f \in \mathcal{A}_1(\mathcal{D}, M)$.

It was proved in [DT2] that for a general dictionary $\mathcal{D}$ the Pure Greedy Algorithm provides the following estimate

$$\|f - G_m(f,\mathcal{D})\| \leq |f|_{\mathcal{A}_1(\mathcal{D})} m^{-1/6}. \tag{8.7}$$

(In this and similar estimates we consider that the inequality holds for all possible choices of $\{G_m\}$.) The paper [DT2] contains also an example of a dictionary $\mathcal{D}$ and an element $f$ such that (see Subsection 8.3 below)

$$\|f - G_m(f,\mathcal{D})\| > \frac{1}{2}|f|_{\mathcal{A}_1(\mathcal{D})} m^{-1/2}, \quad m \geq 4. \tag{8.8}$$

We proved in [KoT2] a new estimate

$$\|f - G_m(f,\mathcal{D})\| \leq 4|f|_{\mathcal{A}_1(\mathcal{D})} m^{-11/62} \tag{8.9}$$

which improves a little the original one (see (8.7)).

E. Livshitz [Li] proved that there exist $\delta > 0$, a dictionary $\mathcal{D}$ and an element $f \in H$, $f \neq 0$, such that

$$\|f - G_m(f, \mathcal{D})\| \geq Cm^{-1/2+\delta}|f|_{\mathcal{A}_1(\mathcal{D})} \tag{8.10}$$

with a positive constant $C$. We developed and refined ideas from [Li] in [T24] and proved the following lower estimate. There exist a dictionary $\mathcal{D}$ and an element $f \in H$, $f \neq 0$, such that

$$\|f - G_m(f, \mathcal{D})\| \geq Cm^{-1/3}|f|_{\mathcal{A}_1(\mathcal{D})} \tag{8.11}$$

with a positive constant $C$.

For the WGA we have the following estimate [T20].

**Theorem 8.7.** *Let $\mathcal{D}$ be an arbitrary dictionary in $H$. Assume $\tau := \{t_k\}_{k=1}^\infty$ is a nonincreasing sequence. Then for $f \in \mathcal{A}_1(\mathcal{D}, M)$ we have*

$$\|f - G_m^\tau(f, \mathcal{D})\| \leq M(1 + \sum_{k=1}^m t_k^2)^{-t_m/2(2+t_m)}. \tag{8.12}$$

In a particular case $\tau = t$, ($t_k = t$, $k = 1, 2, \dots$), (8.12) gives

$$\|f - G_m^t(f, \mathcal{D})\| \leq M(1 + mt^2)^{-t/(4+2t)}, \quad 0 < t \leq 1.$$

This estimate implies the following inequality

$$\|f - G_m^t(f, \mathcal{D})\| \leq C_1(t)m^{-at}|f|_{\mathcal{A}_1(\mathcal{D})}, \quad a < 1/6, \tag{8.13}$$

with the exponent $at$ approaching 0 linearly in $t$. We proved in [T24] that this exponent can not decrease to 0 slower than linearly.

**Theorem 8.8.** *There exists an absolute constant $b > 0$ such that for any $t > 0$ we can find a dictionary $\mathcal{D}_t$ and a function $f_t \in \mathcal{A}_1(\mathcal{D}_t)$ such that*

$$\liminf_{m\to\infty} \|f_t - G_m^t(f_t, \mathcal{D}_t)\| m^{bt}/|f_t|_{\mathcal{A}_1(\mathcal{D}_t)} > 0.$$

We formulate one result for WOGA from [T20]. In the case of OGA this theorem was proved in [DT2].

**Theorem 8.9.** *Let $\mathcal{D}$ be an arbitrary dictionary in $H$. Then for each $f \in \mathcal{A}_1(\mathcal{D}, M)$ we have*

$$\|f - G_m^{o,\tau}(f, \mathcal{D})\| \leq M(1 + \sum_{k=1}^m t_k^2)^{-1/2}.$$

There is one more greedy type algorithm which works well for functions from the convex hull $A_1(\mathcal{D}) := \{f \quad : \quad |f|_{\mathcal{A}_1(\mathcal{D})} \leq 1\}$ of $\mathcal{D}^\pm$, where $\mathcal{D}^\pm := \{\pm g, \quad g \in \mathcal{D}\}$.

There are several modifications of Relaxed Greedy Algorithm (see for instance [B], [DT2]). Before giving the definition of Weak Relaxed Greedy Algorithm (WRGA) we make one remark which helps to motivate the corresponding definition. Assume $G_{m-1} \in A_1(\mathcal{D})$ is an approximant to $f \in A_1(\mathcal{D})$ obtained at the $(m-1)$-th step. The major idea of relaxation in greedy algorithms is to look for an approximant at the $m$-th step of the form $G_m := (1-a)G_{m-1} + ag$, $g \in \mathcal{D}^\pm$, $0 \leq a \leq 1$. This form guarantees that $G_m \in A_1(\mathcal{D})$. We give now the definition of two versions of WRGA.

**Weak Relaxed Greedy Algorithms (WRGA).** *We define* $f_0^{\tau,i} := f$ *and* $G_0^{\tau,i} := 0$ *for* $i = 1, 2$. *Then for each* $m \ge 1$ *we inductively define*

*1).* $\varphi_m^{\tau,1} \in \mathcal{D}^{\pm}$ *is any satisfying*

$$\langle f_{m-1}^{\tau,1}, \varphi_m^{\tau,1} - G_{m-1}^{\tau,1} \rangle \ge t_m \| f_{m-1}^{\tau,1} \|^2 \tag{8.14}$$

*and*

$$\| \varphi_m^{\tau,1} - G_{m-1}^{\tau,1} \| \ge \| f_{m-1}^{\tau,1} \|; \tag{8.15}$$

$\varphi_m^{\tau,2} \in \mathcal{D}^{\pm}$ *is any satisfying*

$$\langle f_{m-1}^{\tau,2}, \varphi_m^{\tau,2} - G_{m-1}^{\tau,2} \rangle \ge t_m \| f_{m-1}^{\tau,2} \|^2. \tag{8.16}$$

*2).*

$$G_m^{\tau,1} := G_m^{\tau,1}(f, \mathcal{D}) := (1 - \alpha_m) G_{m-1}^{\tau,1} + \alpha_m \varphi_m^{\tau,1},$$

$$\alpha_m := \langle f_{m-1}^{\tau,1}, \varphi_m^{\tau,1} - G_{m-1}^{\tau,1} \rangle \| \varphi_m^{\tau,1} - G_{m-1}^{\tau,1} \|^{-2};$$

$$G_m^{\tau,2} := G_m^{\tau,2}(f, \mathcal{D}) := (1 - \beta_m) G_{m-1}^{\tau,2} + \beta_m \varphi_m^{\tau,2},$$

$$\beta_m := t_m (1 + \sum_{k=1}^{m} t_k^2)^{-1} \quad \text{for} \quad m \ge 1.$$

*3).*

$$f_m^{\tau,i} := f - G_m^{\tau,i}, \quad i = 1, 2.$$

We formulate now some theorems on convergence rates of greedy type algorithms WRGA for functions from $\mathcal{A}_1(\mathcal{D}, M)$.

**Theorem 8.10.** *Let* $\mathcal{D}$ *be an arbitrary dictionary in* $H$. *Then for each* $f \in A_1(\mathcal{D})$ *we have*

$$\| f - G_m^{\tau,i}(f, \mathcal{D}) \| \le 2(1 + \sum_{k=1}^{m} t_k^2)^{-1/2}, \quad i = 1, 2. \tag{8.17}$$

We present some results from [T17] on $r$-greedy dictionaries (see Definition 1.3).

**Definition 8.1.** *We say* $\mathcal{D}$ *is a* $\lambda$*-quasiorthogonal dictionary if for any* $n \in \mathcal{N}$ *and any* $g_i \in \mathcal{D}$, $\quad i = 1, \dots, n$, *there exists a collection* $\varphi_j \in \mathcal{D}$, $\quad j = 1, \dots, M$, $\quad M \le N := \lambda n$, *with the properties:*

$$g_i \in X_M := \operatorname{Span}(\varphi_1, \dots, \varphi_M);$$

*and for any* $f \in X_M$ *we have*

$$\max_{1 \le j \le M} |\langle f, \varphi_j \rangle| \ge N^{-1/2} \| f \|.$$

**Theorem 8.11.** *Let a given dictionary $\mathcal{D}$ be $\lambda$-quasiorthogonal and let $0 < r < (2\lambda)^{-1}$ be a real number. Then for any $f$ such that*

$$\sigma_m(f, \mathcal{D}) \leq m^{-r}, \quad m = 1, 2, \dots,$$

*we have*

$$\|f - G_m(f, \mathcal{D})\| \leq C(r, \lambda) m^{-r}, \quad m = 1, 2, \dots.$$

**Remark 8.2.** *It is clear that an orthonormal dictionary is a 1-quasiorthogonal dictionary.*

**Remark 8.3.** *Theorem 8.11 holds if an assumption that $\mathcal{D}$ is $\lambda$-quasiorthogonal is replaced by an assumption that $\mathcal{D}$ is asymptotically $\lambda$-quasiorthogonal. In order to get the definition of asymptotically $\lambda$-quasiorthogonal dictionary we replace $N$ in the Definition 8.1 by $N(n)$ and instead of $N = \lambda n$ we require*

$$\limsup_{n \to \infty} N(n)/n = \lambda.$$

Here are two examples of asymptotically $\lambda$-quasiorthogonal dictionaries.

**Example 8.1.** *The dictionary $\chi := \{f = |J|^{-1/2} \chi_J, \quad J \subset [0, 1)\}$ where $\chi_J$ is the characteristic function of an interval $J$ is an asymptotically 2-quasiorthogonal dictionary.*

**Example 8.2.** *The dictionary $\mathcal{P}(r)$ that consists of functions of the form $f = p\chi_J, \quad \|f\| = 1$, where $p$ is an algebraic polynomial of degree $r - 1$ and $\chi_J$ is the characteristic function of an interval $J$, is asymptotically $2r$-quasiorthogonal.*

**Example 8.3.** *For given $\mu, \gamma \geq 1$ a dictionary $\mathcal{D}$ is called $(\mu, \gamma)$-semistable if for any $g_i \in \mathcal{D}, \quad i = 1, \dots, n$, there exist elements $h_j \in \mathcal{D}, \quad j = 1, \dots, M \leq \mu n$, such that*

$$g_i \in \operatorname{Span}\{h_1, \dots, h_M\}$$

*and for any $c_1, \dots, c_M$ we have*

$$\|\sum_{j=1}^{M} c_j h_j\| \geq \gamma^{-1/2} \left( \sum_{j=1}^{M} c_j^2 \right)^{1/2}.$$

*A $(\mu, \gamma)$-semistable dictionary $\mathcal{D}$ is $\mu\gamma$-quasiorthogonal.*

**8.3. Saturation property of Pure Greedy Algorithm.** We consider in this subsection a generalization of the Pure Greedy Algorithm. Take a fixed number $n \in \mathcal{N}$ and define the basic step of the $n$-dimensional Greedy Algorithm as follows. Find an $n$-term polynomial

$$p_n(f) := p_n(f, \mathcal{D}) = \sum_{n=1}^{n} c_i g_i, \quad g_i \in \mathcal{D}, \quad i = 1, \dots, n,$$

such that (we assume its existence)

$$\|f - p_n(f)\| = \sigma_n(f, \mathcal{D}).$$

Denote

$$G(n, f) := G(n, f, \mathcal{D}) := p_n(f), \qquad R(n, f) := R(n, f, \mathcal{D}) := f - p_n(f).$$

**$n$-dimensional Greedy Algorithm.** *We define $R_0(n,f) := f$ and $G_0(n,f) := 0$. Then, for each $m \geq 1$, we inductively define*

$$\begin{aligned} G_m(n,f) &:= G_m(n,f,\mathcal{D}) := G_{m-1}(n,f) + G(n, R_{m-1}(n,f)) \\ R_m(n,f) &:= R_m(n,f,\mathcal{D}) := f - G_m(n,f) = R(n, R_{m-1}(n,f)). \end{aligned} \tag{8.18}$$

It is clear that a 1-dimensional Greedy Algorithm is a Pure Greedy Algorithm.

For a general dictionary $\mathcal{D}$, and for any $0 < \beta \leq 1$, we define the class of functions

$$\mathcal{A}^o_\beta(\mathcal{D}, M) := \{f \in H : f = \sum_{k\in\Lambda} c_k w_k, \quad w_k \in \mathcal{D},\ |\Lambda| < \infty \text{ and } \sum_{k\in\Lambda} |c_k|^\beta \leq M^\beta\},$$

and we define $\mathcal{A}_\beta(\mathcal{D}, M)$ as the closure (in $H$) of $\mathcal{A}^o_\beta(\mathcal{D}, M)$. Furthermore, we define $\mathcal{A}_\beta(\mathcal{D})$ as the union of the classes $\mathcal{A}_\beta(\mathcal{D}, M)$ over all $M > 0$. For $f \in \mathcal{A}_\beta(\mathcal{D})$, we define the "quasinorm"

$$|f|_{\mathcal{A}_\beta(\mathcal{D})}$$

as the smallest $M$ such that $f \in \mathcal{A}_\beta(\mathcal{D}, M)$. The following general estimate for the error in approximation of functions $f \in \mathcal{A}_\beta(\mathcal{D})$, $\beta \leq 1$, was proved in [DT2].

**Theorem 8.12.** *If $f \in \mathcal{A}_\beta(\mathcal{D})$, $\beta \leq 1$, then for $\alpha := 1/\beta - 1/2$, we have*

$$\sigma_m(f, \mathcal{D}) \leq C|f|_{\mathcal{A}_\beta(\mathcal{D})} m^{-\alpha}, \quad m = 1, 2, \dots. \tag{8.19}$$

*where $C$ depends on $\beta$ if $\beta$ is small.*

In [DT2] we gave an example which showed that replacing a dictionary $\mathcal{B}$ given by an orthogonal basis by a nonorthogonal redundant dictionary $\mathcal{D}$ may damage the efficiency of the Pure Greedy Algorithm. The dictionary $\mathcal{D}$ in our example differs from the dictionary $\mathcal{B}$ by only one suitably chosen element $g$.

Let $\{h_k\}_{k=1}^\infty$ be an orthonormal basis in a Hilbert space $H$ and let $\mathcal{B} = \{h_k\}_{k=1}^\infty$ be the corresponding dictionary. Consider the following element

$$g := Ah_1 + Ah_2 + aA \sum_{k\geq 3} (k(k+1))^{-1/2} h_k$$

with

$$A := (33/89)^{1/2} \quad \text{and} \quad a := (23/11)^{1/2}$$

Then, $\|g\| = 1$. We define the dictionary $\mathcal{D} = \mathcal{B} \cup \{g\}$.

**Theorem 8.13.** *For the function*

$$f = h_1 + h_2$$

*which is in each space $\mathcal{A}_\beta(\mathcal{D})$, $0 < \beta \leq 1$, we have*

$$\|f - G_m(f, \mathcal{D})\| \geq m^{-1/2}, \quad m \geq 4. \tag{8.20}$$

We proved in [T17] that the $n$-dimensional Greedy Algorithm, like the Pure Greedy Algorithm has a saturation property.

**Theorem 8.14.** *For a given n and any orthonormal basis $\{h_k\}_{k=1}^{\infty}$ there exists an element g such that for the dictionary $\mathcal{D} = g \cup \{h_k\}_{k=1}^{\infty}$ there is an element f which has the property: for any $0 < \beta \le 1$*

$$\|f - G_m(n,f)\|/|f|_{\mathcal{A}_\beta(\mathcal{D})} \ge C(\beta) n^{-1/\beta}(m+2)^{-1/2}.$$

**Open Problems.**

**8.1.** Find the order of decay of the sequence

$$\gamma(m) := \sup_{f,\mathcal{D},\{G_m\}} (\|f - G_m(f,\mathcal{D})\| |f|^{-1}_{\mathcal{A}_1(\mathcal{D})}),$$

where sup is taken over all dictionaries $\mathcal{D}$, all elements $f \in \mathcal{A}_1(\mathcal{D}) \setminus \{0\}$ and all possible choices of $\{G_m\}$.

**8.2.** Is there greedy type algorithm realizing (8.19) for $0 < \beta < 1$?

## 9. Greedy Algorithms in Banach spaces

In this section we present some results on greedy approximation with regard to redundant dictionaries in Banach spaces. These results are fragmentary and should be considered as an attempt to understand a role of redundancy and nonlinearity in the general setting for Banach spaces. There is no general results on convergence of $X$-Greedy Algorithm and Dual Greedy Algorithm. Some results about performence of DGA can be found in [Du]. It is proved in [Du] that the assumption that $X$ is a smooth Banach space is a necessary and sufficient condition for the sequence $\{\|R_m^D(f,\mathcal{D})\|_X\}$ to be strictly decreasing for each $f \in X$ and all dictionaries $\mathcal{D}$.

**9.1. Uniformly smooth Banach spaces.** Recently, we proved in [T22] one general convergence result for the generalization of WOGA to Banach spaces. We call this generalization Weak Chebyshev Greedy Algorithm (WCGA). We will use the notation $\mathcal{D}^{\pm} := \{\pm g, \quad g \in \mathcal{D}\}$ here. Let a weakness sequence $\tau = \{t_k\}_{k=1}^{\infty}$, $0 \le t_k \le 1$, be given.

**Weak Chebyshev Greedy Algorithm (WCGA).** *We define $f_0^c := f_0^{c,\tau} := f$. Then for each $m \ge 1$ we inductively define*

*1). $\varphi_m^c := \varphi_m^{c,\tau} \in \mathcal{D}^{\pm}$ is any satisfying*

$$F_{f_{m-1}^c}(\varphi_m^c) \ge t_m \sup_{g \in \mathcal{D}^{\pm}} F_{f_{m-1}^c}(g).$$

*2). Define*

$$\Phi_m := \Phi_m^{\tau} := \mathrm{Span}\{\varphi_j^c\}_{j=1}^m,$$

*and define $G_m^c := G_m^{c,\tau}$ to be the best approximant to $f$ from $\Phi_m$.*

*3). Denote*

$$f_m^c := f_m^{c,\tau} := f - G_m^c.$$

Let us give one more definition of weak greedy type algorithm. We will not present results on it here.

**Weak Dual Greedy Algorithm (WDGA).** *We define* $f_0^D := f_0^{D,\tau} := f$. *Then for each* $m \geq 1$ *we inductively define*

*1).* $\varphi_m^D := \varphi_m^{D,\tau} \in \mathcal{D}^\pm$ *is any satisfying*

$$F_{f_{m-1}^D}(\varphi_m^D) \geq t_m \sup_{g \in \mathcal{D}^\pm} F_{f_{m-1}^D}(g).$$

*2). Define* $a_m$ *as*

$$\|f_{m-1}^D - a_m \varphi_m^D\| = \min_{a \in \mathbb{R}} \|f_{m-1}^D - a_m \varphi_m^D\|.$$

*3). Denote*

$$f_m^D := f_m^{D,\tau} := f_{m-1}^D - a_m \varphi_m^D.$$

We define now the generalization for Banach spaces of the Weak Relaxed Greedy Algorithm studied in [T20] in the case of Hilbert space.

**Weak Relaxed Greedy Algorithm (WRGA).** *We define* $f_0^r := f_0^{r,\tau} := f$ *and* $G_0^r := G_0^{r,\tau} := 0$. *Then for each* $m \geq 1$ *we inductively define*

*1).* $\varphi_m^r := \varphi_m^{r,\tau} \in \mathcal{D}^\pm$ *is any satisfying*

$$F_{f_{m-1}^r}(\varphi_m^r - G_{m-1}^r) \geq t_m \sup_{g \in \mathcal{D}^\pm} F_{f_{m-1}^r}(g - G_{m-1}^r).$$

*2). Find* $0 \leq \lambda_m \leq 1$ *such that*

$$\|f - ((1-\lambda_m)G_{m-1}^r + \lambda_m \varphi_m^r)\| = \inf_{0 \leq \lambda \leq 1} \|f - ((1-\lambda)G_{m-1}^r + \lambda \varphi_m^r)\|$$

*and define*

$$G_m^r := G_m^{r,\tau} := (1-\lambda_m)G_{m-1}^r + \lambda_m \varphi_m^r.$$

*3). Denote*

$$f_m^r := f_m^{r,\tau} := f - G_m^r.$$

**Remark 9.1.** *It follows from the definition of WCGA, WDGA, and WRGA that the sequences* $\{\|f_m^c\|\}$, $\{\|f_m^D\|\}$, *and* $\{\|f_m^r\|\}$ *are nonincreasing sequences.*

The term "weak" in these definitions means that at the step 1). we do not shoot for the optimal element of the dictionary which realizes the corresponding sup but are satisfied with weaker property than being optimal. The obvious reason for this is that we don't know in general that the optimal one exists. Another, practical reason is that the weaker the assumption the easier to satisfy it and therefore easier to realize in practice. The Weak Relaxed Greedy Algorithm provides incremental approximants discussed in [DDGS]. In [DDGS] they also impose weaker assumptions ($\epsilon$-greedy) on an element of the dictionary than being optimal. For instance, for a given sequence $\{\epsilon_n\}_{n=1}^\infty$, $\epsilon_n > 0$, $n = 1, 2, \dots$, they take $0 \leq \alpha_m \leq 1$ and $g_m \in \mathcal{D}$ satisfying

$$\|f - ((1-\alpha_m)G_{m-1} + \alpha_m g_m)\| \leq \inf_{0 \leq \alpha \leq 1, g \in \mathcal{D}} \|f - ((1-\alpha)G_{m-1} + \alpha g)\| + \epsilon_m$$

instead of trying to find optimal ones. Their approach is different from ours.

We discuss in this section the questions of convergence and the rate of convergence for the two above defined methods of approximation: WCGA and WRGA. It is clear that in the case of WRGA the assumption that $f$ belongs to the closure of convex hull of $\mathcal{D}^{\pm}$ is natural. We denote the closure of convex hull of $\mathcal{D}^{\pm}$ by $A(\mathcal{D}) := A_1(\mathcal{D})$. It has been proven in [T20] (see Theorems 8.9 and 8.10 from Section 8) that in the case of Hilbert space both algorithms WCGA and WRGA give the approximation error for the class $A(\mathcal{D})$ of the order

$$(1+\sum_{k=1}^{m} t_k^2)^{-1/2}.$$

We consider here approximation in uniformly smooth Banach spaces. For a Banach space $X$ we define the modulus of smoothness

$$\rho(u) := \sup_{\|x\|=\|y\|=1} (\frac{1}{2}(\|x+uy\|+\|x-uy\|)-1).$$

The uniformly smooth Banach space is the one with the property

$$\lim_{u\to 0} \rho(u)/u = 0.$$

It is easy to see that for any Banach space $X$ its modulus of smoothness $\rho(u)$ is an even convex function satisfying the inequalities

$$\max(0, u-1) \le \rho(u) \le u, \quad u \in (0,\infty). \tag{9.1}$$

It has been established in [DDGS] that the approximation error of an algorithm analogous to our WRGA with $t_k = 1$, $k = 1, 2, \dots$, for the class $A(\mathcal{D})$ can be expressed in terms of modulus of smoothness of Banach space. Namely, if modulus of smoothness $\rho$ of $X$ satisfies the inequality $\rho(u) \le \gamma u^q$, $q > 1$, then the error is of $O(m^{1/q-1})$. It has been proven in [T22] that both algorithms WCGA and WRGA provide approximation for the class $A(\mathcal{D})$ in a Banach space $X$ with modulus of smoothness $\rho(u) \le \gamma u^q$, $1 < q \le 2$, of order

$$(1+\sum_{k=1}^{m} t_k^p)^{-1/p}, \quad p := \frac{q}{q-1}. \tag{9.2}$$

We also proved (see a version of [T22] submitted for publication in Advances of Comp. Math.) that WCGA converges for any $f \in X$ and WRGA converges for any $f \in A(\mathcal{D})$ if $\tau$ satisfies the condition

$$\sum_{m=1}^{\infty} t_m \xi_m(\rho, \tau, \theta) = \infty. \tag{9.3}$$

The sequences $\{\xi_m(\rho,\tau,\theta)\}$ are defined as follows.

**Definition 9.1.** *Let $\rho(u)$ be an even convex function on $(-\infty,\infty)$ with the property: $\rho(2)\ge 1$ and*

$$\lim_{u\to 0}\rho(u)/u=0.$$

*For any $\tau=\{t_k\}_{k=1}^\infty$, $0<t_k\le 1$, and $0<\theta\le 1/2$ we define $\xi_m:=\xi_m(\rho,\tau,\theta)$ as a number $u$ satisfying the equation*

$$\rho(u)=\theta t_m u. \tag{9.4}$$

In a particular case of $\rho(u)\asymp u^q$, $1<q\le 2$, the relation (9.3) is equivalent to

$$\sum_{k=1}^m t_k^p=\infty,\quad p:=\frac{q}{q-1}. \tag{9.5}$$

We gave in [T22] an example which shows that (9.5) is a necessary condition for convergence of WCGA in Banach spaces with modulus of smoothness of power type $q$ for all $\mathcal{D}$ and $f\in X$.

It is well known (see for instance [DDGS], Lemma B.1) that in the case $X=L_p$, $1\le p<\infty$ we have

$$\rho(u)\le\begin{cases} u^p/p & \text{if}\quad 1\le p\le 2,\\ (p-1)u^2/2 & \text{if}\quad 2\le p<\infty.\end{cases} \tag{9.6}$$

It is also known (see [LT], p.63) that for any $X$ with $\dim X=\infty$ one has

$$\rho(u)\ge(1+u^2)^{1/2}-1$$

and for every $X$, $\dim X\ge 2$,

$$\rho(u)\ge Cu^2,\quad C>0.$$

This limits power type modulus of smoothness of nontrivial Banach spaces to the case $1\le q\le 2$. The following theorem gives the rate of convergence of WCGA for $f$ in $A(\mathcal{D})$.

**Theorem 9.1.** *Let $X$ be a uniformly smooth Banach space with the modulus of smoothness $\rho(u)\le\gamma u^q$, $1<q\le 2$. Then for a sequence $\tau:=\{t_k\}_{k=1}^\infty$, $t_k\le 1$, $k=1,2,\dots$, we have for any $f\in A(\mathcal{D})$ that*

$$\|f-G_m^{c,\tau}(f,\mathcal{D})\|\le C(q,\gamma)(1+\sum_{k=1}^m t_k^p)^{-1/p},\quad p:=\frac{q}{q-1},$$

*with a constant $C(q,\gamma)$ which may depend only on $q$ and $\gamma$.*

**9.2. Finite-dimensional spaces.** We discuss some results from [DT3] on $X$-Greedy Algorithms in a particular case of finite-dimensional space $X=\mathbb{R}^n$, equipped with one of standard norms $\ell_p$. The reasons of our concentration on the finite dimensional problems are the following. It is well-known how one can apply the finite dimensional results in studying the smoothness classes. Next, we are interested in understanding an interplay of several parameters including a parameter measuring

the redundancy of a system $\mathcal{D}$. In this subsection it will be more convenient for us to use systems $\mathcal{D}$ that are not necessarily normalized. We note that the definition of $X$-Greedy Algorithm does not depend on normalization of a system.

We use the standard notation $\mathbb{R}^n$ for the $n$-dimensional space of real vectors and the $\ell_p$-norm is defined as follows

$$\|x\|_p := (\sum_{j=1}^n |x_j|^p)^{1/p}, \quad 1 \le p < \infty,$$

$$\|x\|_\infty := \max_j |x_j|.$$

Let $B_p^n$ denote the unit $\ell_p$-ball of $\mathbb{R}^n$.

We give first two theorems from [DT3] about the $m$-term approximation in $\mathbb{R}^n$. In this subsection, we shall consider $m$-term approximation in the $\ell_p$ norm of certain sets $F \subset \mathbb{R}^n$. In Theorem 9.2, we use ideas from [KT1] to give a lower estimate for $m$-term approximation in the $\ell_1$ norm from a general dictionary to general sets $F \subset \mathbb{R}^n$. Lower estimates in the $\ell_1$ norm automatically provide lower estimates in the other $\ell_q$ norms, $q > 1$ (see Corollary 9.1).

We let $\mathrm{Vol}_n(S)$ denote the Euclidean $n$-dimensional volume of the set $S \subset \mathbb{R}^n$. We recall that the volume of the unit ball $B_p^n$, $1 \le p \le \infty$, in $\mathbb{R}^n$ can be estimated by

$$C_1^n n^{-n/p} \le \mathrm{Vol}_n(B_p^n) \le C_2^n n^{-n/p}, \tag{9.7}$$

with $C_1, C_2 > 0$ absolute constants.

**Theorem 9.2.** *If $F \subset B_2^n$ satisfies*

$$\mathrm{Vol}_n\, F \ge K^n\, \mathrm{Vol}_n\, B_2^n,$$

*for some $0 < K \le 1$, then for any dictionary $\mathcal{D}$, $\#\mathcal{D} = N$, we have*

$$\sigma_m(F, \mathcal{D})_1 \ge CK^2 n^{1/2} N^{-\frac{m}{n-m}}, \quad m \le n/2,$$

*with $C > 0$ an absolute constant.*

**Corollary 9.1.** *Let $F$ and $\mathcal{D}$ be as in Theorem 9.2. For any $1 \le q \le \infty$, we have*

$$\sigma_m(F, \mathcal{D})_q \ge CK^2 n^{1/q-1/2} N^{-\frac{m}{n-m}}, \quad m \le n/2.$$

*with $C$ an absolute constant.*

**Corollary 9.2.** *Let $\mathcal{D}$ be as in Theorem 9.2. For any $1 \le p, q \le \infty$, we have*

$$\sigma_m(B_p^n, \mathcal{D})_q \ge C n^{1/q-1/p} N^{-\frac{m}{n-m}}, \quad m \le n/2. \tag{9.8}$$

*with $C$ an absolute constant.*

**Remark 9.2.** *In the case $N = a^n$ and $p = q$, the lower bound in Corollary 9.2 can be replaced by $Ca^{-2m}$*

We shall next consider upper estimates for $\sigma_m(F, \mathcal{D})_p$. We begin with the following simple theorem.

**Theorem 9.3.** *Let $X$ be any $n$-dimensional Banach space and let $B$ be its unit ball. For any $N$ there exists a system $\mathcal{D} \subset X$, $\#\mathcal{D} = N$, such that*

$$\sigma_m(B, \mathcal{D})_X \le \min(1, \epsilon_N^m), \quad \epsilon_N := \frac{2}{N^{1/n} - 1}. \tag{9.9}$$

We consider now the $\ell_p$-Greedy Algorithms, $1 \le p \le \infty$ (see Inrtroduction for the definition). In the case $p = 2$, the $\ell_p$-Greedy Algorithm coincides with the Pure Greedy Algorithm. Then, $G_m^p(x) := G_m^{\ell_p}(x, \mathcal{D})$ is an $m$-term approximation to $x$ from $\mathcal{D}$ which we call the $m$-th greedy approximant. We note that the best approximation to $x \in \mathbb{R}^n$ from $\mathcal{D}$ is not necessarily unique and therefore $G_m^p(x)$ is not necessarily unique. We define

$$\bar{\gamma}_m^p(x, \mathcal{D})_q := \sup \|x - G_m^p(x, \mathcal{D})\|_q$$

where the supremum is taken over all possible resulting $G_m^p(x, \mathcal{D})$. Similarly, we define

$$\underline{\gamma}_m^p(x, \mathcal{D})_q := \inf \|x - G_m^p(x, \mathcal{D})\|_q$$

where the infimum is taken over all possible resulting $G_m^p(x, \mathcal{D})$. Thus, $\bar{\gamma}$ measures the worst possible error over all possible choices of best approximations in the greedy algorithm and $\underline{\gamma}$ represents the best possible error.

More generally, for a class $F \subset \mathbb{R}^n$ we define

$$\bar{\gamma}_m^p(F, \mathcal{D})_q := \sup_{f \in F} \bar{\gamma}_m^p(f, \mathcal{D})_q$$

with a similar definition for $\underline{\gamma}_m^p(F, \mathcal{D})_q$. In upper estimates for greedy approximation we would like to use $\bar{\gamma}$ and for lower estimates $\underline{\gamma}$.

Theorem 9.3 shows that for $p = q$ and for each $a > 1$ there exists a dictionary $\mathcal{D}$, $\#\mathcal{D} = b^n$, $b = 2a + 1$, such that

$$\bar{\gamma}_m^p(B_p^n, \mathcal{D})_p \le a^{-m}.$$

However, the dictionary $\mathcal{D}$ in that theorem is not very natural or easy to describe. This estimate and Remark 9.2 to Corollary 9.2 indicate that systems $\mathcal{D}$ with $\#\mathcal{D}$ of order $C^n$ play an important role in $m$-term approximation in $\mathbb{R}^n$. We proceed now to study a natural family of such systems. We present results from [DT3].

Let $M \ge 3$ be an integer and consider the partition of $[-1, 1]$ into $M$ disjoint intervals $I_i$ of equal length: $|I_i| = 2/M$, $i = 1, \dots, M$. We let $\xi_i$ denote the midpoint of the interval $I_i$, $i = 1, \dots, M$, and $\Xi := \{\xi_i\}_{i=1}^M$. We introduce the system

$$\mathcal{V}_M := \{x \in \mathbb{R}^n : x_j \in \Xi, \ j = 1, \dots n\}.$$

Clearly $|\mathcal{V}_M| = M^n$. We shall study in this section the $\ell_\infty$-Greedy Algorithm for the systems $\mathcal{V}_M$.

**Theorem 9.4.** *For any $1 \le q \le \infty$ we have*

$$\bar{\gamma}_m^\infty(B_\infty^n, \mathcal{V}_M)_q \le n^{1/q} M^{-m}, \quad m = 1, 2, \dots. \tag{9.10}$$

We shall give results about the $\ell_1$ greedy algorithm for the system $\mathcal{V}_3$. We consider this system in detail for the following reasons. It is a simple system which is easy to describe geometrically. Also, it is fairly easy to analyze the approximation properties of this system. Moreoever, it turns out that this system gives geometric order of approximation (see for example Theorem 9.5 and Theorem 9.7) which we know is the best we can expect for general dictionairies (see Corollary 9.2).

**Theorem 9.5.** *We have the estimate*

$$\sigma_m(B_1^n, \mathcal{V}_3)_1 \le \bar{\gamma}_m^1(B_1^n, \mathcal{V}_3)_1 \le (1 - \frac{1}{k+1})^m \tag{9.11}$$

*where* $k := [\log_2(n+1)]$ .

The following lower estimate shows that (9.11) can not be improved by replacing $\log_2(n+1)$ by slower growing function.

**Theorem 9.6.** *Let $n = 4^k - 1$, with $k$ a positive integer. For any $m \le 3k/8$, we have*

$$\underline{\gamma}_m^1(B_1^n, \mathcal{V}_3)_1 \ge 1/2.$$

We want to carry out an analysis similar to the above for the $\ell_2$-Greedy Algorithm (Pure Greedy Algorithm) and the dictionary $\mathcal{V}_3$.

**Theorem 9.7.** *Let $k := [\log_2 n]$. Then,*

$$\bar{\gamma}_m^2(B_2^n, \mathcal{V}_3)_2 \le (1 - \frac{1}{k+1})^{m/2}, \quad m = 1, 2, \dots. \tag{9.12}$$

The following theorem shows that in a certain sense the estimates of Theorem 9.7 cannot be improved.

**Theorem 9.8.** *Let $n = 2^k$ for some positive integer $k$. For any $m \le k/2$, we have*

$$\underline{\gamma}_m^2(B_2^n, \mathcal{V}_3)_2 \ge 1/2.$$

Theorem 9.3 gives the upper estimate for $\sigma_m(B_2^n, \mathcal{D})_2$. In the particular case $\#\mathcal{D} = C^n$, $C > 3$, this theorem guaranties the existence of $\mathcal{D}$ such that

$$\sigma_m(B_2^n, \mathcal{D})_2 \le (\frac{2}{C-1})^m.$$

It is interesting to compare this estimate with the following lower estimate in the problem of selection of optimal basis (see [KT1]). For given $K$ there exists a positive $C(K)$ such that for any set of $S \le K^n$ bases $\mathcal{B}^j$, $j = 1, \dots, S$ in $\mathbb{R}^n$ we have for each $m < n/2$

$$\sup_{f \in B_2^n} \inf_j \sigma_m(f, \mathcal{B}^j)_2 \ge C(K).$$

**Open problems.**

**9.1.** Characterize Banach spaces $X$ such that $X$-Greedy Algorithm converges for all dictionaries $\mathcal{D}$ and each element $f$.

**9.2.** Characterize Banach spaces $X$ such that Dual Greedy Algorithm converges for all dictionaries $\mathcal{D}$ and each element $f$.

**9.3.** Find necessary and sufficient conditions on a weakness sequence $\tau$ to guaranty convergence of Weak Dual Greedy Algorithm in uniformly smooth Banach spaces $X$ with modulus of smoothness of fixed power type $q$, $1 < q \le 2$, ($\rho(u) \le \gamma u^q$) for all dictionaries $\mathcal{D}$ and each element $f \in X$.

**9.4.** Find the correct (in both parameters $n$ and $m$) order of decay of quantities

$$\bar{\gamma}_m^p(B_p^n, \mathcal{V}_3)_p, \quad \underline{\gamma}_m^p(B_p^n, \mathcal{V}_3)_p, \quad p = 1, 2.$$

## 10. Nonlinear $m$-term approximation and $\epsilon$-entropy

In this section, we want to bring out the connection between approximation from a dictionary and $\epsilon$-entropy. We begin with covering numbers $N_\epsilon(F, \ell_p)$ for a set $F \subset \mathbb{R}^n$ and recall their definition. For each $\epsilon > 0$,

$$N_\epsilon(F, l_p) := \min\{N : F \subset \bigcup_{j=1}^{N} B_p^n(y^j, \epsilon)\}$$

with the minimum taken over all sets $\{y_j\}_{j=1}^N$ of points from $\mathbb{R}^n$. Here $B_p^n(y^j, \epsilon)$ denotes the $\ell_p$-ball of radius $\epsilon$ with center $y$. By considering systems $\mathcal{D}$ consisting of the points $y^j$, we find

$$\inf_{\#\mathcal{D}=N_\epsilon(F,\ell_p)} \sigma_1(F, \mathcal{D})_{\ell_p} \le \epsilon. \tag{10.1}$$

In other words, the covering numbers immediately give estimates for 1-term approximation. We can extend the above observation to $m$-term approximation by using the concept of metric entropy. Let $X$ be a linear metric space and for a set $\mathcal{D} \subset X$, let $\mathcal{L}_m(\mathcal{D})$ denote the collection of all linear spaces spanned by $m$ elements of $\mathcal{D}$. For a linear space $L \subset X$, the $\epsilon$-neighborhood $U_\epsilon(L)$ of $L$ is the set of all $x \in X$ which are at a distance not exceeding $\epsilon$ from $L$ (i.e. those $x \in X$ which can be approximated to an error not exceeding $\epsilon$ by the elements of $L$). For any compact set $F \subset X$ and any integers $N, m \ge 1$, we define the $(N, m)$-entropy numbers

$$\epsilon_{N,m}(F, X) := \inf_{\#\mathcal{D}=N} \inf\{\epsilon : F \subset \cup_{L\in\mathcal{L}_m(\mathcal{D})} U_\epsilon(L)\}.$$

We can express $\sigma_m(F, \mathcal{D})$ as

$$\sigma_m(F, \mathcal{D}) = \inf\{\epsilon : F \subset \cup_{L\in\mathcal{L}_m(\mathcal{D})} U_\epsilon(L)\}.$$

It follows therefore that

$$\inf_{\#\mathcal{D}=N} \sigma_m(F, D) = \epsilon_{N,m}(F, X).$$

In other words finding best dictionaries for $m$ term approximation of $F$ is the same as finding sets $\mathcal{D}$ which attain the $(N, m)$-entropy numbers $\epsilon_{N,m}(F, X)$. It is easy to see that $\epsilon_{m,m}(F, X) = d_m(F, X)$. This establishes connection between $(N, m)$-entropy numbers and the Kolmogorov widths.

The present section contains an attempt to generalize the concept of classical Kolmogorov's width in order to be used in estimating best $m$-term approximation. For this purpose we introduce a nonlinear Kolmogorov's $(N, m)$-width:

$$d_m(F, X, N) := \inf_{\Lambda_N, \#\Lambda_N \le N} \sup_{f\in F} \inf_{L\in\Lambda_N} \inf_{g\in L} \|f - g\|_X,$$

where $\Lambda_N$ is a set of at most $N$ $m$-dimensional subspaces $L$. It is clear that

$$d_m(F, X, 1) = d_m(F, X)$$

and

$$d_m(F,X,\binom{N}{m}) \le \epsilon_{N,m}(F,X) \le \sigma_m(F,\mathcal{D})$$

for any $\mathcal{D}$ with $\#\mathcal{D} = N$. The new feature of $d_m(F,X,N)$ is that we allow to choose a subspace $L \in \Lambda_N$ depending on $f \in F$. It is clear that the bigger $N$ the more flexibility we have to approximate $f$. It turns out that from the point of view of our applications the following two cases

$$\text{(I)} \qquad N \asymp K^m,$$

where $K > 1$ is a constant, and

$$\text{(II)} \qquad N \asymp m^{am},$$

where $a > 0$ is a fixed number, play an important role.

We intend to use the $(N,m)$-widths for estimating from below the best $m$-term approximations. There are several general results (see [L], [C]) which give lower estimates of the Kolmogorov widths $d_n(F,X)$ in terms of the entropy numbers $\epsilon_k(F,X)$. In [T16] we generalized the following Carl's (see [C]) inequality: for any $r > 0$ we have

$$\max_{1\le k\le n} k^r \epsilon_k(F,X) \le C(r) \max_{1\le m\le n} m^r d_{m-1}(F,X). \tag{10.2}$$

We denote here for integer $k$

$$\epsilon_k(F,X) := \inf\{\epsilon : \exists f_1,\dots,f_{2^k} \in X : F \subset \cup_{j=1}^{2^k}(f_j + \epsilon B(X))\},$$

where $B(X)$ is the unit ball of Banach space $X$. For noninteger $k$ we set $\epsilon_k(F,X) := \epsilon_{[k]}(F,X)$ where $[k]$ is the integral part of number $k$. It is clear that

$$d_1(F,X,2^n) \le \epsilon_n(F,X).$$

In [T16] we proved the inequality

$$\max_{1\le k\le n} k^r \epsilon_k(F,X) \le C(r,K) \max_{1\le m\le n} m^r d_{m-1}(F,X,K^m), \tag{10.3}$$

where we denote

$$d_0(F,X,N) := \sup_{f\in F} \|f\|_X.$$

This inequality is a generalization of inequality (10.2). In [T16] we also proved the following inequality

$$\max_{1\le k\le n} k^r \epsilon_{(a+r)k\log k}(F,X) \le C \max_{1\le m\le n} m^r d_{m-1}(F,X,m^{am}) \tag{10.4}$$

and gave an example showing that $k\log k$ in this inequality can not be replaced by slower growing function on $k$.

In [T16] we applied inequalities (10.3) and (10.4) for estimating the best $m$-term trigonometric approximation from below. As a corollary to the following version of (10.3) (see Theorem 10.1 below) we gave a new proof (see [DT1]) for the estimate

$$\sigma_m(W^r_\infty,\mathcal{T})_1 \gg m^{-r},$$

where $W^r_\infty$ is a standard Sobolev class (see Section 2) with the restriction imposed in the $L_\infty$-norm.

**Theorem 10.1.** *For any positive constant $K$ we have*

$$\max_{1\le k\le n} k^r \epsilon_k(F,X) \le C(r,K) \max_{1\le m\le n} m^r d_{m-1}(F,X,(Kn/m)^m).$$

We used in [T16] a version of (10.4) to get some new lower estimates of $m$-term trigonometric approximation in the $L_1$-norm of multivariate classes $MW^r_\infty$ of functions with bounded mixed derivative . We proved in [T16] that

$$\sigma_m(MW^r_\infty,\mathcal{T})_1 \gg m^{-r}(\log m)^{r(d-2)}. \tag{10.5}$$

The inequality (10.5) gives a new estimate for small $r$.

The above method can be applied to a general system $\Psi$ instead of trigonometric system $\mathcal{T}$.

Assume a system $\Psi := \{\psi_j\}_{j=1}^\infty$ of elements in $X$ satisfies the condition:

(VP) There exist three positive constants $A_i$, $i=1,2,3$, and a sequence $\{n_k\}_{k=1}^\infty$, $n_{k+1}\le A_1 n_k$, $k=1,2,\dots$ such that there is a sequence of the de la Vallée-Poussin type operators $V_k$ with the properties

$$V_k(\psi_j) = \lambda_{k,j}\psi_j, \tag{10.6}$$

$$\lambda_{k,j}=1 \quad \text{for} \quad j=1,\dots,n_k; \qquad \lambda_{k,j}=0 \quad \text{for} \quad j>A_2 n_k,$$

$$\|V_k\|_{X\to X} \le A_3, \quad k=1,2,\dots \quad . \tag{10.7}$$

**Theorem 10.2.** *Assume that for some $a>0$ and $b\in\mathbb{R}$ we have*

$$\epsilon_m(F,X) \ge C_1 m^{-a}(\log m)^b, \quad m=1,2,\dots \quad .$$

*Then if a system $\Psi$ satisfies the condition (VP) and also satisfies the following condition*

$$E_n(F,\Psi) := \sup_{f\in F}\inf_{c_1,\dots,c_n}\|f-\sum_{j=1}^n c_j\psi_j\|_X \le C_2 n^{-a}(\log n)^b, \quad n=1,2,\dots;$$

*then we have*

$$\sigma_m(F,\Psi)_X \gg m^{-a}(\log m)^b.$$

**Open problem 10.1.** The correct order of the quantity $\sigma_m(MW^r_\infty,\mathcal{T})_1$ is unknown.

## 11. Bilinear Approximation

In this section we discuss one particular case of a dictionary. Denote by $\Pi$ the system of functions of the form $u(x_1)v(x_2)$. It is clear that $\mathcal{T}^2\subset\Pi$. It is also clear that $\Pi$ is a very redundant system. We already mentioned some results for this system in Introduction and in Section 8. All those results concerned approximation in Hilbert space $L_2([0,1]^2)$ and it was convenient for us to normalize elements of $\Pi$ in $L_2$ (what made the system $\Pi$ a dictionary in $L_2([0,1]^2)$). In this section we consider approximation by $\Pi$ in all $L_p$, $1\le p\le\infty$, spaces. In order to make the

system $\Pi$ a dictionary in $L_p$ we need to normalize it in $L_p$. We will denote the normalized in $L_p$ system $\Pi$ by $\Pi_p$. The most results of this section give estimates for best $m$-term approximation. These results do not depend on normalization of $\Pi$ and for convenience in such a case we will use notation $\Pi$ without an index $p$. In this section we concentrate only on approximation of bivariate functions from standard function classes. We note that the bilinear approximation is a well established area now and many estimates are proved in general setting: $f$ is a function of $2d$ variables $x = (x_1, \dots, x_d)$, $y = (y_1, \dots, y_d)$; $\Pi$ is replaced by $\Pi^d := \{u(x)v(y)\}$; $L_p$ is replaced by $L_{p_1,p_2}$, where

$$\|f\|_{p_1,p_2} := \|\|f(\cdot, y)\|_{p_1}\|_{p_2}.$$

The key role in bilinear approximation is played by the Schmidt formula (see Section 8)

$$\sigma_m(f, \Pi)_2 = (\sum_{n=m+1}^{\infty} s_n(J_f)^2)^{1/2}. \tag{11.1}$$

This formula implies in particular for $a > 0$

$$\sigma_m(f, \Pi)_2 \ll m^{-a} \quad \Longleftrightarrow \quad s_n(J_f) \ll m^{-a-1/2}.$$

The following classes are well known and important in studying integral operators. We say that $J_f$ belongs to the Schatten $v$-class $S_v$ if

$$\sum_n s_n(J_f)^v < \infty.$$

The Schmidt formula (11.1) allows us to prove the following result.

**Theorem 11.1.** *For any $v < 2$ we have*

$$J_f \in S_v \quad \Longleftrightarrow \quad \sum_m (\sigma_m(f, \Pi)_2 m^{-1/2})^v < \infty.$$

This theorem is an analog of the following theorem (see [DT2]) for an orthonormal basis $\mathcal{B}$ for a Hilbert space $H$.

**Theorem 11.2.** *For any $\beta < 2$ and any orthonormal basis $\mathcal{B}$ we have*

$$f \in \mathcal{A}_\beta(\mathcal{B}) \quad \Longleftrightarrow \quad \sum_m (\sigma_m(f, \mathcal{B}) m^{-1/2})^\beta < \infty.$$

Theorem 11.2 is the generalization of Stechkin's result [St] that corresponds to $\beta = 1$ in Theorem 11.2. Let us present some general results for approximation in Banach spaces and then get as a corollary error estimates for approximation by $\Pi$ in $L_p$. We remind that $A_1(\mathcal{D})$ is a convex hull of $\mathcal{D}^\pm$. Similarly to the definition of $\mathcal{A}_\beta(\mathcal{D})$ in Subsection 8.3 we define $\mathcal{A}_\beta(\mathcal{D})$ in a Banach space $X$ with a dictionary $\mathcal{D}$. It is easy to derive (see an idea in [DT2, Theorem 3.3]) from Theorem 9.1 the following statement.

**Theorem 11.3.** *Let $X$ be a uniformly smooth Banach space with the modulus of smoothness $\rho(u) \le \gamma u^q$, $1 < q \le 2$. Then for any $f \in \mathcal{A}_\beta(\mathcal{D})$, $0 < \beta \le 1$, we have*

$$\sigma_m(f, \mathcal{D})_X \le C(X, \mathcal{D}) m^{1/q - 1/\beta} |f|_{\mathcal{A}_\beta(\mathcal{D})}.$$

In a particular case $X = L_p$, $1 < p < \infty$, $\mathcal{D} = \Pi_p$, Theorem 11.3 gives the estimate

$$\sigma_m(f, \Pi)_p \le C(p) m^{\max(1/p, 1/2) - 1/\beta} |f|_{\mathcal{A}_\beta(\Pi_p)}. \tag{11.2}$$

This inequality gives the error estimate of best $m$-term approximation in terms of $|f|_{\mathcal{A}_\beta(\Pi_p)}$ which is not well studied for $p \ne 2$. We will present some results on estimates for $\sigma_m(f, \Pi)_p$ in terms of standard Sobolev-Nikol'skii classes. The results from Section 5 (see (5.6) (5.8)) indicate that bilinear approximation of $f(x - y)$ is closely connected with the Kolmogorov widths $d_m(F[f], L_p)$ and best $m$-term approximation of $f$ with regard to the trigonometric system. If $f \in H_q^R$ then $f(x-y) \in NH_q^{(R,R)}$. We get from [T9] that

$$\sigma_m(NH_q^{(R,R)}, \Pi)_p \ll m^{-R + (1/q - \max(1/p, 1/2))_+} \tag{11.3}$$

for $1 \le q \le p \le \infty$ with $R > R(q,p)$, $R(q,p) = 2(1/q - 1/p)$ for $1 \le q \le p \le 2$ and $R(q,p) = 1/q + \max(1/q, 1/2)$ for $p > 2$. Comparing (11.3) with (5.7) we see that the upper estimates for the wider class $NH_q^{(R,R)}$ have the same order as for the class $\{f(x-y), \quad f \in H_q^R\}$. Further results for anisotropic classes $NH_{q_1,q_2}^{(R_1,R_2)}$ and their $2d$-dimensional generalizations can be found in [T9].

In the case $1 \le p \le q \le \infty$ we have

$$\sigma_m(NH_q^{(R,R)}, \Pi)_p \asymp m^{-R}. \tag{11.4}$$

A nontrivial estimate in (11.4) is the lower estimate for $p = 1$, $q = \infty$. This estimate and generalizations of (11.4) are obtained in [T11]. Let us present now results in approximation in the $L_2$-norm for general classes $NH_{q_1,q_2}^{(R_1,R_2)}$ (see [T9] and [T12]). We note that the study of $\sigma_m(NH_{q_1,q_2}^{(R_1,R_2)}, \Pi)_{p_1,p_2}$ is not complete. One of open problems in this area is given in Open problem 11.7. Known results can be found in [T9] and [T12]. Denote $\eta_i := (1/q_i - 1/2)_+$, $i = 1, 2$.

**Theorem 11.4.** *Let $R_1 \le R_2$ and $R_1 > \eta_1$, $R_2 > \eta_2(1 - \eta_1/\eta_2)^{-1}$. Then*

$$\sigma_m(NH_{q_1,q_2}^{(R_1,R_2)}, \Pi)_2 \asymp m^{-R_2(1 - \eta_1/R_1)}, \quad 1 \le q_1, q_2 \le \infty.$$

**Theorem 11.5.** *Let $R_1, R_2$ be as in Theorem 11.4. Then*

$$\sup_{f \in NH_{q_1,q_2}^{(R_1,R_2)}} s_m(J_f) \asymp m^{-R_2(1 - \eta_1/R_1) - 1/2}, \quad 1 \le q_1, q_2 \le \infty.$$

**Theorem 11.6.** *Let $R_1 \ge R_2$, $R_2 > \eta_2$, $R_1 > \eta_1(1 - \eta_2/R_2)^{-1}$. Then*

$$\sigma_m(NH_{q_1,q_2}^{(R_1,R_2)}, \Pi)_2 \asymp m^{-R_1(1 - \eta_2/R_2) + \eta_1 - \eta_2}, \quad 1 \le q_1, q_2 \le \infty.$$

**Theorem 11.7.** *Let $R_1$, $R_2$ be as in Theorem 11.6. Then*

$$\sup_{f\in NH^{(R_1,R_2)}_{q_1,q_2}} s_m(J_f) \asymp m^{-R_1(1-\eta_2/R_2)+\eta_1-\eta_2-1/2}, \quad 1 \le q_1, q_2 \le \infty.$$

We give now some historical remarks on estimating eigenvalues and singular numbers of integral operators. We begin with the following theorem that is a corollary to the Weyl Majorant Theorem (see [GK, p.41]).

**Theorem 11.8.** *Let $A$ be a compact (completely continuous) operator in a Hilbert space $H$. Suppose that*

$$s_n(A) \ll n^{-r}, \quad r > 0.$$

*Then*

$$|\lambda_n(A)| \ll n^{-r}.$$

Fredholm [F] proved that if the kernel $f(x,y)$ is a continuous function and satisfies the condition

$$\sup_{x,y} |f(x,y+t) - f(x,y)| \le C|t|^\alpha, \quad 0 < \alpha \le 1,$$

then for an arbitrary $\rho > 2/(2\alpha+1)$ the series

$$\sum_{j=1}^{\infty} |\lambda_j(J_f)|^\rho < \infty$$

converges.

Starting with that article, smoothness conditions with respect to one variable were imposed on the kernel. Weyl [We] proved the estimate

$$\lambda_n(J_f) = o(n^{-r-1/2})$$

under the condition that the kernel $f(x,y)$ is symmetric and continuous and that $\partial^r f/\partial x^r$ is continuous. Let us introduce some more notation. Define $NH^{(R,0)}_{q_1,q_2}$ as follows: $f(x,y)$ belongs to this class if for all $y \in \mathbb{T}$ the function $f(\cdot,y)$ of $x$ belongs to the class $H^R_{q_1}B(y)$, and $B(y)$ is such that $\|B(y)\|_{q_2} \le 1$. We use here the following notation. For a function class $F$ and a number $B > 0$ we define $FB := \{f : f/B \in F\}$.

Hille and Tamarkin [HT] achieved significant progress. They proved, in particular, that for $1 < q \le 2$ and $R \ge 1$

$$\sup_{f\in NH^{(R,0)}_{q,q'}} |\lambda_n(J_f)| \ll n^{-R-1+1/q}(\log n)^R, \quad q' = q/(q-1),$$

and they conjectured that the extra logarithmic factor can be removed or even replaced by a logarithmic factor with a negative power.

The next important step was taken by Smithies [Sm]. He proved the estimate

$$\sup_{f\in NH^{(R,0)}_{q,2}} s_n(J_f) \ll n^{-R-1+1/q}, \quad 1 < q \le 2, \quad R > 1/q - 1/2. \tag{11.5}$$

Of later results we mention those of Gel'fond and M.G. Krein (see [GK, Ch.III, S10.4]), Birman and Solomyak [BS], and Cochran [Co].

We proved in [T9] the following estimate

$$\sigma_m(NH^{(R,0)}_{q_1,q_2},\Pi)_{p_1,p_2} \asymp m^{-R+(1/q_1-\max(1/2,1/p_1))},$$

for $1 \le q_1 \le p_1 \le \infty$, $1 \le q_2 = p_2 \le \infty$ and $R > r(q_1,p_1)$. We denote here $r(q,p) := (1/q-1/p)_+$ for $1 \le q \le p \le 2$ or $1 \le p \le q \le \infty$ and $r(q,p) := \max(1/2,1/q)$ otherwise. This inequality implies in particular that (11.5) holds also for $q = 1$.

We discuss now an application of bilinear approximation to the theory of widths. As we know the starting point of this theory is a function class, say, the function class $W^r_q$. This function class can be associated with one function - the Bernoulli kernel $F_r(x-y)$ with

$$F_r(t) := 2\sum_{k=1}^{\infty} k^{-r}\cos(kt - r\pi/2).$$

We have

$$W^r_q = \{f \quad : \quad f(x) = \hat f(0) + (2\pi)^{-1}\int_0^{2\pi} f_r(x-y)\varphi(y)dy, \quad \|\varphi\|_q \le 1\}.$$

In the development of approximation by trigonometric polynomials it was understood that the rate of decay of $E_n(f)$ of individual functions, say $E_n(F_r)$, is governed by smoothness properties of the function. It turned out that we have similar phenomenon on the much more general level.

For a function $g \in L_1(\mathbb{T}^2)$ define a function class

$$W^g_q := \{f \quad : \quad f(x) = (2\pi)^{-1}\int_0^{2\pi} g(x,y)\varphi(y)dy, \quad \|\varphi\|_q \le 1\}.$$

We proved in [T7] that $F_\rho(x-y)$ is a typical representative of the following class of functions. Denote $MH^{r_1,r_2}_1B$ the class of functions $g(x,y)$ such that $\|g\|_1 < \infty$,

$$\int_0^{2\pi} g(x,y)dx = \int_0^{2\pi} g(x,y)dy = 0$$

(this condition is imposed only for convenience), and

$$\|\Delta^l_{t_1,t_2} g(x,y)\|_1 \le B|t_1|^{r_1}|t_2|^{r_2}, \quad r_1,r_2 > 0, \quad l := \max([r_1],[r_2]) + 1,$$

where $\Delta^l_{t_1,t_2}$ denotes the operator of the mixed difference of order $l$ in each variable with step $t_1$ in $x$ and step $t_2$ in $y$. We remark that the function $F_\rho(x-y)$ belongs to $MH^{r_1,r_2}_1B$ for any $r_1,r_2$ such that $r_1 + r_2 = \rho$. We proved in [T7] the following statement.

**Theorem 11.9.** *For all $1 \le q,p \le \infty$ we have*

$$\sup_{g\in MH^{r_1,r_2}_1} d_m(W^g_q, L_p) \asymp d_m(W^{r_1+r_2}_q, L_p)$$

*for $r_1 > 1$, $r_2 > 1 + \max(1/q,1/2)$ for $2 \le q < p \le \infty$ or $1 \le q < 2 < p \le \infty$ and $r_2 > 1$ otherwise.*

**Open problems.**

**11.1.** Find necessary and sufficient conditions on a weakness sequence $\tau$ to guaranty convergence of Weak Greedy Algorithm with regard to $\Pi_2$ for each $f \in L_2$.

**11.2.** Does $L_p$-Greedy Algorithm with regard to $\Pi_p$ converge for each $f \in L_p$, $1 < p < \infty$?

**11.3.** Does Dual Greedy Algorithm with regard to $\Pi_p$ converge for each $f \in L_p$, $1 < p < \infty$?

**11.4.** If the answer to Problem 11.3 is "yes" then find necessary and sufficient conditions on a weakness sequence $\tau$ to guaranty convergence of Weak Dual Greedy Algorithm with regard to $\Pi_p$ for each $f \in L_p$.

**11.5.** Find necessary and sufficient conditions on a weakness sequence $\tau$ to guaranty convergence of Weak Chebyshev Greedy Algorithm with regard to $\Pi_p$ for each $f \in L_p$.

**11.6.** Let $R_N$ be the Rudin-Shapiro polynomials (see Section 4). Prove that

$$\sigma_m(R_N(x-y), \Pi)_1 \gg N^{1/2}.$$

**11.7.** Find the order of the sequence

$$\sigma_m(NH_1^{(R_1,R_2)}, \Pi)_{p_1,\infty}, \quad m = 1, 2, \dots \tag{11.6}$$

in the case $R_1 < R_2$, $2 < p_1 \le \infty$.

*Comment.* In the case $R_1 \ge R_2$ the order of (11.6) is known (see [T9]).

**11.8.** Study efficiency of Pure Greedy Algorithm ($L_2$-Greedy Algorithm) with regard to $\Pi_2$ for approximation of function classes $NH_{q_1,q_2}^{(R_1,R_2)}$ in the $L_{p_1,p_2}$-norm.

## 12. Ridge Approximation

This section similarly to Section 11 is devoted to approximation of functions of two variables. The results discussed here may be seen as one more (in addition to Section 11) example in the development of the following general approach in multivariate approximation. Approximate functions of several variables by univariate functions. This idea is interesting from theoretical point of view and also looks reasonable from computational point of view. There is a number of different realizations of this approach in approximation theory. We mention some of them for illustration. We begin with the simplest one. S.N. Bernstein (see [Be]) suggested to study the following type of approximation to a continuous periodic function $f(x, y)$ on two variables

$$E_{n,\infty}(f) := \inf_{\{c_k(y)\}} \|f(x,y) - \sum_{|k| \le n} c_k(y) e^{ikx}\| \tag{12.1}$$

in the uniform norm $\|\cdot\|$. The approximant in (12.1) is a linear combination of products of univariate functions. The Bernstein setting of the problem (12.1) is a variant of the classical problem of bilinear approximation which was discussed in Section 11. The important feature of the problem of bilinear approximation is that the approximating system $\{u(x)v(y)\}_{u,v \in L_2}$ is highly redundant. However, as we have seen in Section 11 the redundancy did not hinder the development of nice theory to solve the problem of best bilinear approximation in the $L_2$-norm. What really allowed to do it is the structure of the system. In this section we

discuss approximation by a redundant system with quite different structure. We approximate by linear combinations of ridge functions, i.e. functions $G(x), x \in \mathbb{R}^2$, which can be represented in the form

$$G(x) = g((x, e)) \tag{12.2}$$

where $g$ is a univariate function and its argument $(x, e)$ is the scalar product of $x$ and a unit vector $e \in \mathbb{R}^2$. We denote the set of functions of the form (12.2) by $\mathcal{R}$ and call it the system of ridge functions. The above mentioned approximation (approximation by ridge functions) also uses univariate functions and the system $\mathcal{R}$ of all ridge functions is highly redundant. Unlike the bilinear approximation problem we do not have a theory which provides (describes) the solution to the problem of best ridge approximation. In this section we confine ourselves to the case of functions of two variables and approximate only in Hilbert space $L_2$. We note that approximation by ridge functions got much attention recently for the following two reasons. The first is that a ridge function can be interpreted as a plane wave. This means that the problem of ridge approximation can be seen as a problem of representation of a general wave by plane waves. The second reason is that ridge approximation proved to be useful in neural networks approximation (see [DOP]).

There are some general results on approximation by linear combinations of elements of a redundant system in Hilbert space (see Theorem 11.3). These results are expressed in terms of the $\mathcal{A}_\beta(\mathcal{D})$-quasinorm determined by a dictionary $\mathcal{D}$. Let $D := \{(x_1, x_2) : x_1^2 + x_2^2 \le 1\}$ be the unit disk and $L_p(D)$, $1 \le p < \infty$, denote the Banach space with the norm

$$\|f\|_p := \|f\|_{L_p(D)} := (\frac{1}{\pi}\int_D |f(x)|^p dx)^{1/p}.$$

From this point on we denote by $\mathcal{R}_p$ the dictionary for $L_p(D)$ which consists of elements of the system $\mathcal{R}$ normalized in $L_p(D)$. Similarly to the bilinear approximation we use the notation $\mathcal{R}$ instead of $\mathcal{R}_p$ when we talk about best $m$-term approximations. In a particular case $X = L_p(D)$, $1 < p < \infty$, $\mathcal{D} = \mathcal{R}_p$ Theorem 11.3 gives the estimate

$$\sigma_m(f, \mathcal{R})_p \le C(p) m^{\max(1/p, 1/2) - 1/\beta} |f|_{\mathcal{A}_\beta(\mathcal{R}_p)}. \tag{12.3}$$

This inequality gives the error estimate of best $m$-term approximation in terms of $|f|_{\mathcal{A}_\beta(\mathcal{R}_p)}$ which is not well studied. In order to use this general result we need to varify that a given function $f$ can be approximated by functions which have special representation (see definition of $\mathcal{A}_\beta(\mathcal{D})$), what in turn could be a nontrivial problem. We will present some results on estimates for $\sigma_m(f, \mathcal{R})_p$ in terms of standard classes of functions. In this section we deal with the function class which is defined in a way standard for constructive approximation. We define the class of functions $H_p^r(D)$ using the classical means of approximation, namely, algebraic polynomials. Let $\mathcal{P}(n, 2)$ denote the set of algebraic polynomials

$$\sum_{k+l \le n-1} c_{k,l} x_1^k x_2^l$$

of total degree $n-1$. Denote by $H_p^r(D)$, $r>0$, the set of all functions $f \in L_p(D)$ which can be repsented in the form

$$f = \sum_{n=1}^{\infty} p_n, \qquad p_n \in \mathcal{P}(2^n, 2), \quad n = 1, 2, \dots,$$

with $p_n$ satisfying the inequalities

$$\|p_n\|_p \le 2^{-rn}.$$

The following result (see [LS]) gives the upper estimates for $\sigma_m(H_p^r(D), \mathcal{R})_p$ automatically.

**Theorem 12.1.** *For any algebraic polynomial $p \in \mathcal{P}(N,2)$ there exist $N$ univariate polynomials $g^j$, $j = 0, \dots, N-1$, of degree $N-1$ with the following property*

$$p(x) = \sum_{j=0}^{N-1} g^j((x, e_j^N)), \tag{12.4}$$

*where $e_j^N := (\cos\frac{j\pi}{N}, \sin\frac{j\pi}{N})$.*

This gives the estimate

$$\sigma_m(H_p^r(D), \mathcal{R})_p \le C(r) m^{-r}. \tag{12.5}$$

It turned out that in the case $p=2$ the estimate (12.5) is sharp:

$$\sigma_m(H_2^r(D), \mathcal{R})_2 \ge C(r) m^{-r}. \tag{12.6}$$

The first result in this direction a little weaker version of (12.6) was obtained in [T13]. The estimate (12.6) was proved in [Ma]. The estimate (12.6) also follows from the relation

$$\sigma_m(f, \mathcal{R})_2 \ge C \inf_{p \in \mathcal{P}(3m,2)} \|f - p\|_2 \tag{12.7}$$

established in [O3] for radial functions $f$, $f(x_1, x_2) = h((x_1^2 + x_2^2)^{1/2})$.

We proved recently (see [MOT]) that the estimate (12.5) in the case $p=2$ can be realized by PGA

$$\sup_{f \in H_2^r} \|f - G_m(f, \mathcal{R}_2)\|_2 \le C(r) m^{-r}. \tag{12.8}$$

Let us make some comments on (12.8). First of all this estimate shows that PGA with regard to $\mathcal{R}_2$ is not saturated. Moreover, combining (12.8) with (12.7) we get that for radial functions $f$ such that

$$\sigma_m(f, \mathcal{R})_2 \le C(r) m^{-r} \tag{12.9}$$

we have

$$\|f - G_m(f, \mathcal{R}_2)\|_2 \le C(r) m^{-r}.$$

This is a weaker analog of the $r$-greedy property for $\mathcal{R}_2$.

**Open problems.**

**12.1.** Find necessary and sufficient conditions on a weakness sequence $\tau$ to guaranty convergence of Weak Greedy Algorithm with regard to $\mathcal{R}_2$ for each $f \in L_2$.

**12.2.** Does $L_p$-Greedy Algorithm with regard to $\mathcal{R}_p$ converge for each $f \in L_p$, $1 < p < \infty$?

**12.3.** Does Dual Greedy Algorithm with regard to $\mathcal{R}_p$ converge for each $f \in L_p$, $1 < p < \infty$?

**12.4.** If the answer to Problem 12.3 is "yes" then find necessary and sufficient conditions on a weakness sequence $\tau$ to guaranty convergence of Weak Dual Greedy Algorithm with regard to $\mathcal{R}_p$ for each $f \in L_p$.

**12.5.** Find necessary and sufficient conditions on a weakness sequence $\tau$ to guaranty convergence of Weak Chebyshev Greedy Algorithm with regard to $\mathcal{R}_p$ for each $f \in L_p$.

**12.6.** Find the order of the quantity

$$\sup_{f \in \mathcal{A}_1(\mathcal{R}_2)} \|f - G_m(f, \mathcal{R}_2)\|_{L_2(\mathcal{D})}.$$

**12.7.** Could the estimate (12.5) for $1 < p < \infty$ be realized by WCGA with $\tau = \{t\}$, $0 < t \le 1$?

## References


[AT] A.V. Andrianov and V.N. Temlyakov, *Best $m$-term approximation of functions from classes $MW^r_q$*, Approximation Theory IX (to appear).

[B] Andrew R. Barron, *Universal approximation byounds for superposition of $n$ sigmoidal functions*, IEEE Transactions on Information Theory **39** (1993), 930– 945.

[Ba] B.M. Baishanski, *Approximation by polynomials of given length*, Illinois J. Math. **27** (1983), 449–458.

[Bab] K.I. Babenko, *Some problems in approximation theory and numerical analysis*, Russian Math. Surveys **40** (1985), 1–30.

[Be] S.N. Bernstein, *On the best approximation of functions of several variables by means of polynomials of trigonometric sums*, Trudy Mat. Inst. Steklov **38** (1951), 24–29.

[BS] M.Sh. Birman and M.Z. Solomyak, *Estimates of singular numbers of integral operators*, Uspekhi Mat. Nauk **32** (1977), 17–84; English transl. in Russian Math. Surveys **32** (1977).

[C] B. Carl, *Entropy numbers, s-numbers, and eigenvalue problems*, J. Funct. Anal. **41** (1981), 290–306.

[CDH] A. Cohen, R.A. DeVore, and R. Hochmuth, *Restricted Nonlinear Approximation*, Constructive Approx. **16** (2000), 85–113.

[CF] A. Cordoba and P. Fernandez, *Convergence and divergence of decreasing rearranged Fourier series*, SIAM, I. Math. Anal. **29** (1998), 1129–1139.

[Co] J.A. Cochran, *Composite integral operators and nuclearity*, Ark. Mat. **15** (1977), 215–222.

[D] R.A. DeVore, *Nonlinear approximation*, Acta Numerica (1998), 51–150.

[DDGS] C. Darken, M. Donahue, L. Gurvits, E. Sontag, *Rate of convex approximation in non-Hilbert spaces*, Constructive Approximation **13** (1997), 187–220.

[DJP] R. DeVore, B. Jawerth, and V. Popov, *Compression of wavelet decompositions*, American Journal of Mathematics **114** (1992), 737–785.

[DKT] R.A. DeVore, S.V. Konyagin, and V.N. Temlyakov, *Hyperbolic wavelet approximation*, Constructive Approximation **14** (1998), 1–26.

[DKTe] S. Dilworth, D. Kutzarova, and V. Temlyakov, *Convergence of some Greedy Algorithms in Banach spaces*, Manuscript (2001).

[DL] R.A. DeVore and G.G. Lorenz, *Constructive Approximation*, Springer-Verlag, Berlin, 1993.

[DMA] G. Davis, S. Mallat, and M. Avellaneda, *Adaptive greedy approximations*, Constructive Approximation **13** (1997), 57–98.

[Do1] D.L. Donoho, *Unconditional bases are optimal bases for data compression and for statistical estimation*, Appl. Comput. Harmon. Anal. **1** (1993), 100–115.

[Do2] D.L. Donoho, *CART and Best-Ortho-Basis: A Connection*, Preprint (1995), 1–45.

[DOP] R.A. DeVore, K.I. Oskolkov, and P.P. Petrushev, *Approximation by feedforward neural networks*, Annals of Numerical Mathematics **4** (1997), 261–287.

[DP] R.A. DeVore and V.A. Popov, *Interpolation spaces and non-linear approximation*, Lecture Notes in Mathematics **1302** (1988), 191–205.

[DT1] R.A. DeVore and V.N. Temlyakov, *Nonlinear approximation by trigonometric sums*, J. Fourier Analysis and Applications **2** (1995), 29–48.

[DT2] R.A. DeVore and V.N. Temlyakov, *Some remarks on Greedy Algorithms*, Advances in computational Mathematics **5** (1996), 173–187.

[DT3] R.A. DeVore and V.N. Temlyakov, *Nonlinear approximation in finite-dimensional spaces*, J. Complexity **13** (1997), 489–508.

[Du] V.V. Dubinin, *Greedy Algorithms and Applications*, Ph.D. Thesis, University of South Carolina, 1997.

[F] I. Fredholm, *Sur une classe d'équations fonctionalles*, Acta Math. **27** (1903), 365–390.

[FJ] M. Frazier and B. Jawerth, *A discrete transform and decomposition of distribution spaces*, J. Functional Analysis **93** (1990), 34-170.

[FS] J.H. Friedman and W. Stuetzle, *Projection pursuit regression*, J. Amer. Statist. Assoc. **76** (1981), 817–823.

[GK] I.C. Gohberg and M.G. Krein, *Introduction to the Theory of Linear Nonselfadjoint Operators in Hilbert space*, AMS, Providence, Rhode Island 02904, 1969.

[H] P.J. Huber, *Projection Pursuit*, The Annals of Statistics **13** (1985), 435–475.

[HT] E. Hille and J.D. Tamarkin, *On the characteristic values of linear integral equations*, Acta Math. **57** (1931), 1–76.

[I] R.S. Ismagilov, *Widths of sets in normed linear spaces and the approximation of functions by trigonometric polynomials*, Uspekhi Mat. Nauk **29** (1974), 161–178; English transl. in Russian Math. Surveys **29** (1974).

[J1] L. Jones, *On a conjecture of Huber concerning the convergence of projection pursuit regression*, The Annals of Statistics **15** (1987), 880–882.

[J2] L. Jones, *A simple lemma on greedy approximation in Hilbert space and convergence rates for projection pursuit regression and neural network training*, The Annals of Statistics **20** (1992), 608–613.

[K1] T.W. Körner, *Divergence of decreasing rearraged Fourier series*, Annals of Mathematics **144** (1996), 167–180.

[K2] T.W. Körner, *Decreasing rearraged Fourier series*, The J. Fourier Analysis and Applications **5** (1999), 1–19.

[Ka1] B.S. Kashin, *Widths of certain finite-dimensional sets and classes of smooth functions*, Izv. Akad. Nauk SSSR, Ser.Mat. **41** (1977), 334–351; English transl. in Math. USSR IZV. **11** (1977).

[Ka2] B.S. Kashin, *On approximation properties of complete orthonormal systems*, Trudy Mat. Inst. Steklov **172** (1985), 187–191; English transl. in Proc. Steklov Inst. Math. **1987**, no. 3, 207–211.

[Kah] J.P. Kahane, *Series De Fourier Absolument Convergentes*, Springer-Verlag, Berlin, 1976.

[KoT1] S.V. Konyagin and V.N. Temlyakov, *A remark on greedy approximation in Banach spaces*, East J. on Approx. **5** (1999), 1–15.

[KoT2] S.V. Konyagin and V.N. Temlyakov, *Rate of convergence of Pure Greedy Algorithm*, East J. on Approx. **5** (1999), 493–499.

[KT1] B. S. Kashin and V. N. Temlyakov, *On best $m$-terms approximations and the entropy of sets in the space $L^1$*, Math. Notes **56** (1994), 57–86.

[KT2] B.S. Kashin and V.N. Temlyakov, *On estimating approximative characteristics of classes of functions with bounded mixed derivative*, Math. Notes **58** (1995), 922–925.

[KS] B.S. Kashin and A.A. Saakyan, *Orthogonal Series*, American Math. Soc., Providence, R.I., 1989.

[L] G.G. Lorentz, *Metric entropy and approximation*, Bull. Amer. Math. Soc. **72** (1966), 903–937.

[Le] H. Lebesgue, *Sur les intégrales singuliéres*, Ann. Fac. Sci. Univ. Toulouse (3) **1** (1909), 25–117.

[LGM] G.G. Lorentz, M.v. Golitschek, Yu. Makovoz, *Constructive Approximation*, Springer-Verlag, Berlin, 1996.

[Li] E.D. Livshitz, *On the rate of convergence of greedy algorithm*, Manuscript (2000).

[LS] B.F. Logan and L.A. Shepp, *Optimal reconstruction of a function from its projections*, Duke Math. Journal **42** (1975), 645–659.

[LT] J. Lindenstrauss and L. Tzafriri, *Classical Banach Spaces I*, Springer-Verlag, Berlin, 1977.

[LTe] E.D. Livshitz and V.N. Temlyakov, *On convergence of Weak Greedy Algorithms,*, IMI-Preprint **13** (2000), 1–9.

[Ma] V.E. Maiorov, *On best approximation by ridge functions*, Preprint, Technion, Izrael **27** (1998).

[MOT] V.E. Maiorov, K.I. Oskolkov, and V.N. Temlyakov, *Gridge Approximations and Radon Compass*, IMI-Preprint **9** (2000), 1–20.

[N1] S. N. Nikol'skii, *Approximation of functions of several variables and embedding theorems*, (English translation of Russian published by "Nauka", Moskow, 1969), Springer-Verlag.

[N2] S.M. Nikol'skii, *On interpolation and best approximation of differentiable periodic functions by trigonometric polynomials*, Izv. Akad. Nauk SSSR Ser. Mat. **10** (1946), 393–410.

[O1] K.I. Oskolkov, *An estimate in the approximation of continuous functions by subsequences of Fourier sums*, Proc. Steklov Inst. Math. **134** (1975), 273–288.

[O2] K.I. Oskolkov, *On the Lebesgue inequality in the uniform metric and on a set of full measure*, Mat. Zametki **18** (1975), 515–526.

[Os] P. Oswald, *Greedy algorithms and best $m$-term approximation with respect to biorthogonal systems*, Preprint (2000), 1–22.

[P] P. Petrushev, *Direct and converse theorems for spline and rational approximation and Besov spaces*, Lecture Notes in Mathematics **1302** (1988), 363–377.

[Pi] G. Pisier, *The volume of convex bodies and Banach space geometry*, Cambridge University Press, 1989.

[Po] V.V. Pospelov, *On approximation of multivariate functions by products of univariate functions*, Inst. Applied Math. Acad. Sciences USSR, Preprint No. 32 (1978), 1–76.

[RW] L. Rejtö and G.G. Walter, *Remarks on projection pursuit regression and density estimation*, Stochastic Analysis and Application **10** (1992), 213–222.

[S] E. Schmidt, *Zur Theorie der linearen und nichtlinearen Integralgleichungen. I*, Math. Annalen **63** (1906-1907), 433–476.

[Sm] F. Smithies, *The eigen-values and singular values of integral equations*, Proc. London Math. Soc. (2) **43** (1937), 255–279.

[St] S. B. Stechkin, *On absolute convergence of orthogonal series*, Dok. Akad. Nauk SSSR **102** (1955), 37-40.

[T1] V.N. Temlyakov, *On the Asymptotic Behavior of Best Approximations of Continuous Functions*, Soviet Math. Dokl. **17** (1976), 739-743.

[T2] V.N. Temlyakov, *Asymptotic Behavior of Best Approximations of Continuous Functions*, Math USSR Izvestija **11** (1977), 551-569.

[T3] V.N. Temlyakov, *Approximation of Continuous Functions by Trigonometric Polynomials*, Proc. Steklov Inst. Math. (1983), 213-228.

[T4] V.N. Temlyakov, *Widths of Some Classes of Functions of Several Variables*, Soviet Math Dokl. **26** (1982), 619-622.

[T5] V.N. Temlyakov, *On the Asymptotic Behavior of Best Approximations of Inidividual Functions*, Proc. Steklov Inst. Math. **3** (1987), 341-352.

[T6] V.N. Temlyakov, *On Best Bilinear Approximations of Periodic Functions of Several Variables*, Soviet Math. Dokl. **33** (1986), 96-99.

[T7] V.N. Temlyakov, *Estimates of the Best Bilinear Approximations of Functions of Two Variables and Some of Their Applications*, Math. USSR Sb. **62** (1989), 95-109.

[T8] V.N. Temlyakov, *Approximation by elements of a finite dimensional subspaces of functions from various Sobolev or Nikol'skii spaces*, Matem. Zametki **43** (1988), 770–786; English transl. in Math. Notes **43** (1988), 444–454.

[T9] V.N. Temlyakov, *Bilinear Approximation and Applications*, Proc. Steklov Inst. Math. **3** (1990), 221-248.

[T10] V.N. Temlyakov, *On Estimates of Approximation Numbers and Best Bilinear Approximation*, Constructive Approximation **8** (1992), 23-33.

[T11] V.N. Temlyakov, *Estimates of Best Bilinear Approximations of Functions and Approximation Numbers of Integral Operators*, Math. Notes **51** (1992), 510-517.

[T12] V.N. Temlyakov, *Bilinear Approximation and Related Questions*, Proc. Steklov Inst. of Math. **4** (1993), 245-265.

[T13] V.N. Temlyakov, *On approximation by ridge functions*, Preprint, University of South Carolina, USA (1996), 1–12.

[T14] V.N. Temlyakov, *The best $m$-term approximation and Greedy Algorithms*, Advances in Comp. Math. **8** (1998), 249–265.

[T15] V.N. Temlyakov, *Nonlinear $m$-term approximation with regard to the multivariate Haar system*, East J. Approx. **4** (1998), 87–106.

[T16] V.N. Temlyakov, *Nonlinear Kolmogorov's widths*, Matem. Zametki **63** (1998), 891–902.

[T17] V.N. Temlyakov, *Greedy Algorithms and $m$-term Approximation With Regard to Redundant Dictionaries*, J. Approx. Theory **98** (1999), 117–145.

[T18] V.N. Temlyakov, *Greedy algorithms with regard to multivariate systems with special structure*, Constructive Approximation **16** (2000), 399–425.

[T19] V.N. Temlyakov, *Greedy Algorithm and $m$-term Trigonometric Approximation*, Constructive Approx. **14** (1998), 569–587.

[T20] V.N. Temlyakov, *Weak greedy algorithms*, Advances in Comp. Math. **12** (2000), 213–227.

[T21] V.N. Temlyakov, *Universal bases and greedy algorithms*, IMI-Preprint series **8** (1999), 1–20.

[T22] V.N. Temlyakov, *Greedy algorithms in Banach spaces*, IMI-Preprint series **10** (2000), 1–11.

[T23] V.N. Temlyakov, *A criterion for convergence of Weak Greedy Algorithms*, IMI-Preprint series **21** (2000), 1–10.

[T24] V.N. Temlyakov, *Two lower estimates in greedy approximation*, IMI-Preprint series **07** (2001), 1–12.

[Ta] T. Tao, *On the almost everywhere convergence of wavelet summation methods*, Applied and Computational Harmonic Analysis **3** (1996), 384–387.

[Te1] V. Temlyakov, *Approximation of functions with bounded mixed derivative*, Proc. Steklov Institute (1989, Issue 1).

[Te2] V.N. Temlyakov, *Approximation of periodic functions*, Nova Science Publishers, Inc., New York, 1993.

[Ti] V.M. Tikhomirov, *Widths of sets in function spaces and the theory of best approximations*, Uspekhi Mat. Nauk **15** (1960), 81–120; English transl. in Russian Math. Surveys **15** (1960).

[Tim] A.F. Timan, *Theory of approximation of functions of a real variable*, Mac. Millan, New York, 1963.

[W1] P. Wojtaszczyk, *On unconditional polynomial bases in $L_p$ and Bergman spaces*, Constructive Approximation **13** (1997), 1–15.

[W2] P. Wojtaszczyk, *Greedy algorithms for general systems*, J. Approx. Theory **107** (2000), 293–314.

[We] H. Weyl, *Das asymptotische Verteilungsgesetz der Eiegenwerte linearer partieller Differentialgleichungen*, Math. Ann. **71** (1911), 441–479.

[Z] A. Zygmund, *Trigonometric series*, University Press, Cambridge, 1959.